

„SAPIENTIA”

MONICA ALBU

Construirea și utilizarea testelor psihologice

015 7252

CASA DE EDITURĂ „ATLAS-CLUSIUM” SRL

(Director general: VALENTIN TAȘCU)

EDITURA „CLUSIUM”

(Director: NICOLAE MOCANU)

ROMÂNIA, 3400 CLUJ-NAPOCA, Piața Unirii 1

telefax +40-64-196940

© Editura CLUSIUM 1998

CLUSIUM

1998

BIBLIOTECĂ
COT 155.28.443
INVENTAR 13.7.91

Coperta: Lucian ANDREI

PREFATĂ

Rolul important al matematicii în toate domeniile vieții și ale științei a fost recunoscut de foarte multă vreme. Leonardo da Vinci susținea că "nu există nici o certitudine în științele în care nu este posibilă aplicarea vreuneia din științele matematice", iar A. Einstein afirma că "matematica le oferă științelor naturii o doză de siguranță, pe care acestea nu ar putea să o obțină altfel".

Statistica, la rândul ei, are o aplicabilitate generală în domeniul științific. "În toate cercetările științifice de bază, statisticienii sunt cei care elaborează planurile de experiență sau de observații și tot la ei se recurge pentru a analiza rezultatele, pentru a evalua constatările și pentru a separa faptele clar demonstrate de cele care mai cer confirmarea", observa Ronald Fischer.

La fel ca și celelalte discipline științifice, psihologia folosește modele matematice și statistice la explorarea relațiilor dintre variabile și la testarea ipotezelor. Dar, datorită complexității și ambiguității unor aspecte ale comportamentului uman, cu ajutorul procedeele de calcul obișnuite este dificilă realizarea unor inferențe valide și precise. Din acest motiv, au fost create modele noi, specifice științelor sociale, cum sunt teoria răspunsului la itemi, folosită la construirea testelor psihologice, și teoria variabilelor latente (care cuprinde analiza factorială și modelele cu ecuații structurale), utilă pentru studierea relațiilor dintre variabilele măsurate prin testele psihologice.

Lucrarea de față are ca principal obiectiv familiarizarea cititorului cu metodele matematice și statistice care se folosesc la construirea testelor psihologice și la interpretarea scorurilor testelor. Alături de procedeele de calcul uzuale sunt prezentate metode noi, bazate pe teoria aproximării și pe teoria clasificării. Acestea îmbină ideile teoriei clasice a testelor cu cele ale teoriei răspunsului la itemi.

Căutându-se răspuns la întrebarea "Ce înseamnă un test bun?" s-a realizat o sinteză a ideilor cuprinse în lucrări clasice și moderne din domeniul teoriei testelor psihologice. Este propus și un punct de vedere nou asupra teoriei testelor, bazat pe teoria clasificării: testul psihologic este privit ca un instrument de ierarhizare și de clasare a subiecților, iar construirea unui test psihologic este asimilată operației de creare a unui clasament prin agregarea clasamentelor produse de itemi.

Lucrarea oferă, în plus, informații despre condițiile de utilizare a testelor psihologice. Sunt discutate aspecte privind cine, când și cum poate folosi un test psihologic.

Două anexe conțin definițiile noțiunilor matematice și statistice care intervin în metodele de calcul prezentate. Ele sunt însoțite de exemple.

Toți algoritmi proprii propuși au fost verificați pe loturi de subiecți. La administrarea testelor în acest scop și la interpretarea, sub aspect psihologic, a rezultatelor obținute am primit ajutor din partea domnului prof. dr. Horia Pitariu.

Cartea se adresează în primul rând studenților de la facultățile de psihologie, dar ea le poate fi utilă și psihologilor care, în activitatea lor curentă, fac uz de teste psihologice.

CUPRINS

PREFAȚĂ	5
1. TEST ȘI TESTARE PSIHOLAGICĂ	13
1.1. CE ESTE UN TEST PSIHOLAGIC ?	13
1.2. UTILIZĂRI ALE TESTELOR PSIHOLOGICE	18
1.3. CONDIȚII DE UTILIZARE A TESTELOR PSIHOLOGICE	20
2. CÂTEVA NOȚIUNI CARE INTERVIN ÎN TEORIA TESTELOR PSIHOLOGICE	26
2.1. CONSTRUCTUL PSIHOLAGIC	26
2.2. DOMENIUL DE CONȚINUT AL UNUI TEST	29
2.3. GRUPELE CONTRASTANTE	30
2.3.1. Definiția grupelor contrastante	30
2.3.2. Formarea grupelor contrastante	32
2.4. CRITERIUL	35
2.5. VALIDITATEA DE ASPECT A TESTELOR	40
2.6. VARIABILE MODERATOARE, VARIABILE MEDIATOARE	40
3. VECHI ȘI NOU ÎN TESTAREA PSIHOLAGICĂ	44
3.1. SCURT ISTORIC AL TESTELOR PSIHOLOGICE	44
3.2. CÂTEVA TEORII ALE PERSONALITĂȚII ȘI IMPACTUL LOR ASUPRA PSIHOMETRIEI	47
3.2.1. Definiții	47
3.2.2. Câteva teorii ale personalității	50
3.3. EVALUARE PSIHOMETRICĂ SAU EVALUARE COMPORTAMENTALĂ ?	55
3.4. TENDINȚE NOI ÎN STUDIUL PERSONALITĂȚII: MODELUL BIG FIVE ȘI MODELUL CIRCUMPLEX	59
3.4.1. Modelul Big Five	59
3.4.2. Modelul circumplex	63
3.5. UTILIZAREA CALCULATORILOR ÎN TESTAREA PSIHOLAGICĂ	65
4. CLASIFICĂRI ALE TESTELOR PSIHOLOGICE	69
4.1. CLASIFICĂRI ÎN FUNCȚIE DE CONȚINUTUL TESTULUI ..	69
4.2. CLASIFICĂRI ÎN FUNCȚIE DE MODUL DE ADMINISTRARE A TESTULUI	80
4.3. CLASIFICĂRI ÎN FUNCȚIE DE MODUL DE COTARE A TESTULUI ȘI DE INTERPRETARE A SCORURILOR	82
5. CONSTRUIREA ȘI COTAREA ITEMILOR TESTELOR PSIHOLOGICE ..	85

5.1. TIPURI DE ITEMI ÎN TESTELE PSIHOLOGICE	85
5.2. SUGESTII PENTRU SCRIEREA ITEMILOR	92
5.3. CALCULUL SCORULUI UNUI ITEM	95
6. PRINCIPALELE TEORII ASUPRA TESTELOR PSIHOLOGICE	101
6.1. TEORIA CLASICĂ A TESTELOR	101
6.1.1. Scor real, scor observat	101
6.1.2. Instrumente de măsură echivalente, instrumente de măsură paralele	104
6.1.3. Eroarea standard de măsură	107
6.2. TEORIA RĂSPUNSULUI LA ITEMI	108
6.2.1. Variabile latente, variabile observate	108
6.2.2. Funcția caracteristică a unui item	112
6.2.3. Utilizarea rezultatelor la un test	117
6.3. TEORIA STĂRILOR ȘI A TRĂSĂTURILOR LATENTE	121
6.3.1. Ideea de bază a teoriei stărilor și a trăsăturilor latente	121
6.3.2. Scopul teoriei stărilor și a trăsăturilor latente	122
6.3.3. Variabile utilizate de teoria stărilor și a trăsăturilor latente	123
6.3.4. Modelul de bază al teoriei stărilor și a trăsăturilor latente	126
6.3.5. Coeficienții utilizați de teoria stărilor și a trăsăturilor latente	127
6.3.6. Câteva modele ale teoriei stărilor și a trăsăturilor latente	128
7. SURSE DE EROARE ÎN TESTAREA PSIHOLOGICĂ	130
7.1. NOȚIUNEA DE EROARE	130
7.2. FACTORI CARE INFLUENȚEAZĂ REZULTATELE TESTĂRII PSIHOLOGICE	130
8. FIDELITATEA TESTELOR PSIHOLOGICE	136
8.1. DEFINIȚIA FIDELITĂȚII TESTELOR	136
8.2. ROLUL COEFICIENTULUI DE FIDELITATE	138
8.3. FORMULE DE CALCUL UTILIZATE PENTRU EVALUAREA FIDELITĂȚII	140
8.4. METODE PENTRU EVALUAREA FIDELITĂȚII TESTELOR	147
8.4.1. Metoda test-retest	148
8.4.2. Metoda analizei consistenței interne	150
8.4.3. Metoda formelor paralele	152
8.4.4. Coeficienți de fidelitate inter-evaluatori	155

8.5. OBSERVAȚII ASUPRA METODELOR FOLOSITE PENTRU ESTIMAREA FIDELITĂȚII	156
8.6. FACTORI CARE INTERVIN ÎN ESTIMAREA COEFICIENTILOR DE FIDELITATE	157
8.7. INTERPRETAREA COEFICIENTILOR DE FIDELITATE	158
8.8. FIDELITATEA DIFERENȚEI SCORURILOR ÎNTRE TEST ȘI RETEST	159
9. VALIDITATEA TESTELOR PSIHOLOGICE	161
9.1. DEFINIȚIA VALIDITĂȚII	161
9.2. VALIDITATEA TESTULUI RELATIVĂ LA CONSTRUCTUL MĂSURAT DE ACESTA	165
9.3. VALIDITATEA TESTULUI RELATIVĂ LA CONȚINUTUL SĂU	170
9.4. VALIDITATEA TESTULUI RELATIVĂ LA UN CRITERIU	173
9.4.1. Metode de cercetare a validității relative la criteriu ..	173
9.4.2. Problema restrângerii mulțimii	176
9.4.3. Interpretarea coeficienților de validitate	177
9.4.4. Factori care afectează valoarea coeficienților de validitate	178
9.5. CÂTEVA IDEI IMPORTANTE DESPRE VALIDITATE A TESTELOR	181
10. DIFICULTATEA ȘI SENSIBILITATEA UNUI TEST PSIHOLOGIC	183
10.1. DIFICULTATEA UNUI TEST PSIHOLOGIC	183
10.2. SENSIBILITATEA UNUI TEST PSIHOLOGIC	186
10.3. INFLUENȚA CARACTERISTICILOR ITEMILOR ASUPRA DIFICULTĂȚII ȘI A SENSIBILITĂȚII TESTULUI	190
11. INFORMAȚII DESPRE TESTELE PSIHOLOGICE	195
11.1. MANUALELE TESTELOR PSIHOLOGICE	195
11.2. INFORMAȚII NECESARE PENTRU ALEGEREA UNUI TEST PSIHOLOGIC	195
12. CONSTRUIREA TESTELOR PSIHOLOGICE	198
12.1. ALGORITMUL GENERAL DE CONSTRUIRE A TESTELOR	198
12.1.1. Prezentarea pașilor algoritmului de construire a unui test	198
12.1.2. Observații asupra metodelor de construire a testelor	200
12.2. DEFINIREA TESTULUI	201
12.3. CREAREA BĂNCII DE ITEMI	205
12.4. CONSTRUIREA UNEI VARIANTE A TESTULUI	209
12.5. ADMINISTRAREA TESTULUI	211

12.6. ANALIZA REZULTATELOR TESTĂRII	211
12.6.1. Analiza de itemi	212
12.6.1.1. Indicele de dificultate al unui item	213
12.6.1.2. Analiza răspunsurilor incorecte	216
12.6.1.3. Validitatea itemilor	219
12.6.1.4. Indicele de discriminare al unui item	221
12.6.1.5. Indicele de fidelitate al itemului	225
12.6.1.6. Observații asupra analizei de itemi	225
12.6.2. Metode de selectare a itemilor pentru construirea unui test psihologic	226
12.7. EXPERIMENTAREA SUPLIMENTARĂ A TESTULUI	234
12.8. TRADUCEREA ȘI ADAPTAREA UNUI TEST	236
13. INTERPRETAREA SCORURILOR TESTELOR PSIHOLOGICE	237
13.1. PRINCIPII GENERALE	237
13.2. INTERPRETAREA SCORURILOR LA TESTELE DE CUNOȘTINȚE INFORMATIVE	239
13.3. INTERPRETAREA SCORURILOR LA TESTELE CENTRATE PE CRITERIU	240
13.4. INTERPRETAREA SCORURILOR LA TESTELE NORMATIVE	242
13.4.1. Normele testelor	242
13.4.2. Etalonarea testelor	243
13.4.3. Clasele echivalente	247
13.4.4. Vârstele echivalente	249
13.4.5. Normele de vârstă mintală	249
13.4.6. Cotele standard z și cotele standardizate	250
13.4.7. Rangurile centile, decile și cuartile	251
13.4.8. Clasele scării normalizate	256
13.5. INTERPRETAREA SCORURILOR LA CHESTIONARELE DE PERSONALITATE	261
13.5.1. Descrierea procedurii obișnuit	261
13.5.2. O metodă de interpretare a scorurilor în cazul unei singure scale	263
13.5.3. O metodă de interpretare a scorurilor în cazul unei perechi de scale	266
13.5.4. Avantajele celor două metode de interpretare a scorurilor	267
13.5.5. Exemple	268
14. PROFILUL PSIHOLOGIC	273

15. O ABORDARE A TEORIEI TESTELOR PSIHOLOGICE BAZATĂ PE TEORIA CLASIFICĂRII	282
15.1. TESTUL PSIHOLOGIC CA INSTRUMENT DE CLASIFICARE	282
15.2. CONSTRUIREA TESTELOR PSIHOLOGICE	283
15.3. VALIDITATEA UNUI TEST	289
15.3.1. Validitatea relativă la construct	289
15.3.2. Validitatea relativă la criteriu	295
ANEXA I. NOȚIUNI DE TEORIA CLASIFICĂRII	302
1.1. RELAȚII CARE INTERVIN ÎN CLASIFICARE	302
1.2. PARTIȚII	304
1.3. NUMEROTARE A UNEI PREORDINI TOTALE	306
1.4. CLASAMENTE	308
1.5. COMPARAREA A DOUĂ PARTIȚII	310
1.6. COMPARAREA A DOUĂ CLASAMENTE	313
1.7. AGREGAREA CLASAMENTELOR	316
1.8. OPERAȚIILE CLASIFICĂRII	318
ANEXA II. CÂTEVA NOȚIUNI MATEMATICE ȘI STATISTICE CARE INTERVIN ÎN TEORIA TESTELOR PSIHOLOGICE	319
II.1. DEFINIȚIA VARIABILEI	319
II.2. VARIABILE DEPENDENTE, VARIABILE INDEPENDENTE	319
II.3. SCALE DE MĂSURĂ	320
II.4. DISTANȚE	328
II.5. COTE STANDARD, COTE STANDARDIZATE	330
BIBLIOGRAFIE	333

CAPITOLUL 1

TEST ȘI TESTARE PSIHOLOGICĂ

1.1. CE ESTE UN TEST PSIHOLOGIC ?

Testul este un instrument al metodei experimentale, folosit cu precădere în investigațiile cu caracter aplicativ ale psihologiei (testul psihologic), pedagogiei (testul pedagogic, testul școlar), sociologiei (testul sociometric) și biologiei (testul fiziologic).

Testul psihologic reprezintă, alături de observații, interviuri, scale de evaluare, tehnici proiective etc., o sursă de informare utilizată pentru evaluarea comportamentului uman și a proceselor mintale.

În privința definiției testului psihologic, nu s-a ajuns la un consens, fiecare autor oprindu-se asupra altor aspecte sau reținând alte caracteristici. De exemplu:

- Testul psihologic este o măsurătoare obiectivă și standardizată a unui eșantion de comportamente (Anastasi, 1976).
 - Testul psihologic este o situație standardizată care generează un comportament semnificativ (Klausnitzer, 1992).
 - Testul psihologic constă dintr-o probă, mai frecvent dintr-o serie de probe, construite în scopul stabilirii prezenței (sau a absenței) unui aspect psihic, a particularităților de comportare sau a gradului de dezvoltare psihică (Roșca, 1972).
 - Testul psihologic este o probă, utilizată mai ales în psihologia diferențială, care permite descrierea comportamentului unui subiect într-o situație definită precis ("instrucțiunile" testului), prin raportare la comportamentul unei grupe de subiecți fixată, plasată în aceeași situație. Această descriere se face de obicei sub formă numerică (Grand dictionnaire de la psychologie, 1994).
 - Un test psihologic este o procedură standardizată prin care se formează un eșantion de comportamente care va fi descris prin categorii sau scoruri. În plus, multe teste au norme sau standarde care fac posibilă utilizarea rezultatelor la predicția altor comportamente, mai importante (Gregory, 1992).
 - Un test psihologic este definit prin următoarele caracteristici: materiale și proceduri standardizate; motivație optimă; înregistrare imediată; notare obiectivă; norme potrivite și validitate verificată (Dahlström, 1993).
- Fără a încerca să formulăm o definiție, vom prezenta elementele pe care le considerăm esențiale pentru înțelegerea noțiunii de "test psihologic".

A. Testele psihologice servesc fie la *măsurarea unor atribute psihice* (ca, de exemplu, atenția), fie la *predicția unor comportamente* (ca, de pildă, succesul profesional).

B. Un test psihologic constă din una sau mai multe probe. Fiecare probă *determină o reacție comportamentală* a subiectului (cum ar fi, de pildă, reproducerea unui șir de numere) sau îi cere acestuia să facă aprecieri asupra unui comportament al său (de exemplu, să indice frecvența întârzierilor la serviciu ori intensitatea efortului depus pentru îndeplinirea la timp a sarcinilor de serviciu). Răspunsurile sunt evaluate, conform unor reguli fixate, obținându-se una sau mai multe valori numerice, denumite *scoruri*.

Un element al testului psihologic constituit dintr-o situație stimul care solicită un răspuns ce poate fi evaluat separat de restul testului se numește *item*. Un item poate consta, de exemplu, dintr-o întrebare însoțită de mai multe răspunsuri posibile, dintr-un desen care trebuie completat sau dintr-o listă de termeni care trebuie ordonați după un criteriu specificat. Răspunsului dat de subiect la un item i se atribuie un scor.

Un set de itemi care se referă la același aspect (un atribut psihic pe care testul îl măsoară sau un comportament pe care testul îl prezice) formează o *scală*. Pentru fiecare scală se calculează un scor, pe baza scorurilor itemilor componenți, conform unei reguli stabilite de autorul testului. De exemplu, scorul unei scale dintr-un test de cunoștințe poate fi egal cu suma scorurilor itemilor rezolvați corect sau cu scorul celui mai dificil item rezolvat corect sau cu scorul celui mai ușor item rezolvat greșit.

Nu există restricții privind numărul de itemi care formează o scală sau numărul de scale care intră în componența unui test psihologic. Multe chestionare de personalitate cuprind zeci de scale, fiecare conținând mai mulți itemi. În schimb, alte teste (cum este, de pildă, Testul de capacități organizatorice al lui A. Huth) sunt alcătuite dintr-o singură scală, cu un singur item.

C. Pentru a nu necesita un timp de administrare prea lung, nici un test psihologic nu evaluează toate comportamentele care ar putea fi utilizate la definirea sau la măsurarea unui anumit atribut, ci doar un *eșantion de comportamente* din această mulțime, ales astfel încât să fie *reprezentativ* pentru toate comportamentele mulțimii respective care ar putea fi observate în afara situației de testare.

De exemplu, dacă s-ar dori să se evalueze deprinderea copiilor din clasa întâi de a face calcule aritmetice, procedeul corect de lucru ar fi să se urmărească toate calculele efectuate de fiecare subiect: la școală, la rezolvarea temelor, la joacă, la cumpărături etc. Evident, acest lucru este imposibil. S-ar putea, însă, construi un test alegând exerciții care să solicite efectuarea operațiilor aritmetice cu numere formate din una sau două cifre. Pentru ca

eșantionul să fie reprezentativ, exercițiile cuprinse în test trebuie să conțină toate cele patru operații aritmetice.

Nu este necesar ca itemii unui test construit în scopul predicției unui comportament să descrie exact comportamentul respectiv. De altfel, ar fi imposibil ca un test utilizat la selecția profesională, deci înainte de ocuparea unui loc de muncă sau a unui loc la un curs de instruire pentru o anumită profesie, să se refere la comportamentele care sunt specifice profesiei respective. De exemplu, nu li se poate cere persoanelor care se prezintă la selecția pentru școala de șoferi să conducă o mașină pe șosea. *Testul psihologic măsoară comportamente prin care se pot face inferențe asupra altor comportamente*.

D. În etapa de construcție a testului psihologic se stabilesc și regulile care trebuiesc respectate la utilizarea acestuia, pentru a se asigura o funcționare corectă a testului ca instrument de măsură. Aceste reguli, care sunt prezentate în manualul testului, se referă la:

- conținutul testului (care sunt stimulii prezentați pentru a provoca reacțiile comportamentale evaluate de test);
- condițiile de administrare (instrucțiunile date subiectului în legătură cu sarcina ce trebuie executată);
- tehnica de evaluare a răspunsurilor (modul de cotare a reacțiilor, cotare care să nu permită decât o intervenție redusă a subiectivității examinatorului și, în consecință, să realizeze un acord cât mai mare între examinatori);
- modul de interpretare a rezultatelor obținute de un subiect.

O probă sau o serie de probe formează un test psihologic numai dacă este însoțită de *regulile de utilizare*, adică dacă este *standardizată*.

Vom explica în continuare importanța respectării regulilor de utilizare a testului.

• Majoritatea testelor au o structură fixă: tuturor subiecților li se administrează aceiași itemi, în aceeași ordine. Dacă testul conține atât itemi ușori cât și dificili, iar ordinea itemilor nu ar fi fixă, atunci comportarea la test a subiecților ar putea fi influențată de poziția itemilor dificili: prezența acestora la început poate fi obositoare și chiar descurajatoare, repercutându-se negativ asupra modului de răspuns la ceilalți itemi.

Există și teste al căror conținut se stabilește în timpul administrării lor, în funcție de răspunsurile date de subiect.

• Condițiile de administrare a testului afectează comportamentul celor testați. O persoană are șanse mai mici să răspundă corect la itemii unui test de atenție atunci când acesta este administrat într-o încăpere de trecere, decât dacă testarea se face într-o cameră liniștită.

În majoritatea cazurilor modul în care este administrat testul depinde și de competența examinatorului (de gradul de informare al acestuia relativ la test, de experiența pe care o are în administrarea testelor, de grija pe care o manifestă față de examinați).

Dacă testul este administrat colectiv și sunt urmate întocmai instrucțiunile din manualul testului, li se poate asigura tuturor subiecților o similaritate a condițiilor de testare. În schimb, la testarea individuală personalitatea examinatorului are un rol mult mai important, fiind posibil ca aceeași persoană, cu ocazia a două testări efectuate de examinatori diferiți, să aibă comportamente total diferite.

- Cele mai multe teste psihologice au reguli fixe pentru calculul scorurilor, formulate astfel încât scorurile obținute de subiecți să nu depindă de persoana care le calculează. În această categorie intră, de exemplu, testele cu alegere multiplă, întrucât pentru ele este stabilit, încă din faza de construcție a lor, punctajul care se acordă fiecărui răspuns propus, la fiecare item. Există și teste pentru care scorul reflectă aprecierea subiectivă a celui care îl calculează. Un exemplu îl constituie testele de creativitate care cer să se indice cât mai multe posibilități de utilizare pentru un anumit obiect, iar punctajul acordat fiecărui răspuns este stabilit de examinator, în funcție de cât de neobișnuită i se pare utilizarea propusă.

Pentru a cerceta concordanța aprecierilor făcute de mai mulți evaluatori, D. Stark și R. C. Elliot (Vernon, 1965) au trimis câte o copie a unei lucrări de geometrie tuturor profesorilor de matematică de la 16 școli, cerându-le să îi acorde un punctaj în conformitate cu modul lor obișnuit de notare. Rezultatele obținute au variat între 28 și 92 de puncte. Autorii experimentului au încercat să explice variabilitatea mare a notelor prin existența mai multor aspecte care sunt luate în considerare la stabilirea punctajului: corectitudinea soluției, originalitatea sa, claritatea redactării lucrării etc. În absența unor instrucțiuni pentru acordarea notei, fiecare examinator atribuie alte ponderi acestor aspecte.

Testele eseu, care solicită formularea unui răspuns mai amplu la o anumită problemă, intră în aceeași categorie, a testelor cu cotare subiectivă.

- Scorurile obținute de o persoană, doar prin valoarea lor, nu aduc nici o informație. Chiar știind că scorurile unei scale care măsoară memoria pot fi cuprinse între 0 și 30, este imposibil de apreciat dacă un rezultat egal cu 20 indică o memorie obișnuită sau una foarte bună.

Interpretarea unui scor individual se poate face numai prin compararea sa, fie cu rezultatele obținute de alte persoane care fac parte din aceeași populație ca și subiectul, fie cu anumite praguri, indicate de autorul testului, care delimitează mai multe categorii. De exemplu, în cazul testelor de cunoștințe utilizate la concursurile de admitere, care furnizează un punctaj cuprins între 1

și 10, se fixează un singur prag, de obicei egal cu 5, iar subiectul este considerat "admis" dacă scorul său este mai mare decât pragul, și "respins", în caz contrar.

Uneori, pe baza scorurilor obținute la scalele unui test, se identifică o categorie în care este potrivit să fie repartizată persoana examinată. Categoria poate reprezenta un tip de temperament, o boală psihică, un program de pregătire profesională etc.

Modul de interpretare a rezultatelor este stabilit de autorul testului și este prezentat în manualul testului.

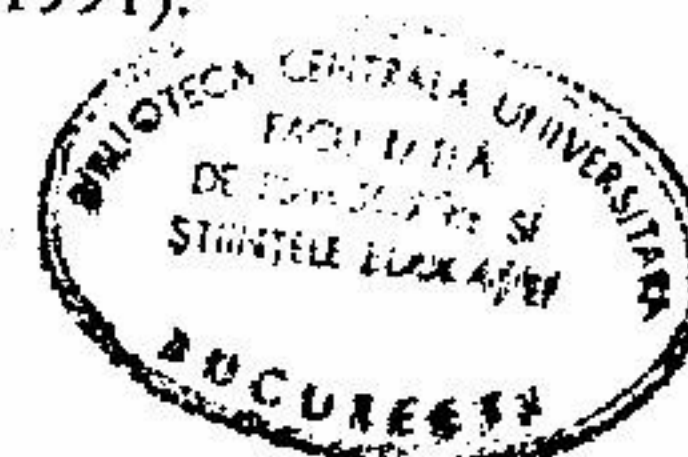
E. Se spune că un test este *obiectiv* dacă el permite să se măsoare fără ambiguitate capacitățile unei persoane (Klausnitzer, 1992). Pentru a fi obiectiv, testul trebuie să îndeplinească trei condiții:

- Rezultatele să nu fie dependente de nivelul de pregătire al subiectului în domeniul testării psihologice, adică persoana examinată să nu poată ghici cum se interpretează răspunsurile la itemi și, prin urmare, să nu poată denatura rezultatele (să nu poată da acele răspunsuri care generează despre sine impresia pe care o dorește).

- Fiecare răspuns posibil la un item să fie interpretat la fel (ca fiind corect sau greșit, favorabil sau nefavorabil pentru aspectul măsurat etc.) de mai mulți experți care analizează testul. De exemplu, la itemul "Mă simt bine într-o societate veselă" mai mulți psihologi să fie de acord că răspunsul "Adevărat" corespunde persoanelor extravertite, iar răspunsul "Fals" celor introvertite.

- Scorul obținut de subiect să nu fie influențat de personalitatea examinatorului, adică testul să fie cu cotare obiectivă. Această condiție nu este îndeplinită de multe teste, în special de cele proiective și de testele de cunoștințe ai căror itemi sunt cu răspunsuri deschise.

F. O trăsătură comună tuturor testelor psihologice este precizia lor limitată. Sunt extrem de rare cazurile când un test furnizează o măsură exactă a unei variabile care are efecte importante asupra comportamentului uman. Din acest motiv, nici deciziile care se iau pe baza lor nu sunt totdeauna corecte. De exemplu, unii psihologi susțin că, din cauza lipsei de consistență a comportamentului de la o situație la alta, coeficienții de corelație liniară dintre scorurile unui test de personalitate și un comportament pe care acesta îl prezice nu depășesc valoarea de 0.3 sau 0.4 (W. Mischel și R. E. Nisbett, după Funder, 1983). Dar, nu există alte metode de măsurare, mai precise. Prin urmare, *testele psihologice reprezintă cea mai bună, cea mai corectă și cea mai precisă tehnologie disponibilă pentru a lua decizii importante asupra indivizilor* (Murphy, Davidshofer, 1991).



1.2. UTILIZĂRI ALE TESTELOR PSIHOLOGICE

Testele psihologice sunt folosite, în principal, pentru a stabili un diagnostic psihologic, prin evaluarea comportamentului, a abilităților mintale și a altor caracteristici de personalitate, pentru a face aprecieri și predicții referitoare la subiecți și pentru a lua decizii asupra persoanelor. În general, testele predictive și cele pentru decizie se utilizează în selecția profesională, iar cele diagnostice servesc la măsurarea unor caracteristici psihice, fie pentru aprecierea efectelor unui program de instruire, fie pentru cunoașterea individului de către psihologul clinician.

Predicția se referă la estimarea performanței viitoare a individului, de exemplu la un anumit loc de muncă.

Într-un sens restrâns, însă, chiar și diagnosticul situației prezente, de pildă al retardării mintale sau al dezechilibrului emoțional, implică o predicție asupra a ceea ce va face individul în situații diferite de cea actuală. Din acest motiv, se poate afirma că toate testele psihologice sunt eșantioane de comportament pe baza cărora se pot face predicții asupra altor comportamente (Anastasi, 1976).

Pe de altă parte, luarea unei decizii corecte se bazează de obicei pe efectuarea unei prognoze asupra consecințelor deciziei. De exemplu, într-o selecție de personal decizia de încadrare pe un post se ia în urma comparării performanțelor trecute și actuale ale indivizilor cu un set de condiții, considerate a fi absolut necesare pentru îndeplinirea muncii respective. Scopul deciziei este de a identifica acele persoane care vor realiza cele mai bune performanțe pe postul pe care vor fi încadrate. Variabilele care servesc ca bază pentru decizie și asupra cărora sunt fixate condițiile eliminatorii trebuie să aibă valoare predictivă.

În funcție de scopul urmărit prin administrarea testelor, se pot distinge următoarele categorii de utilizări ale acestora (Gregory, 1992):

A. Clasificarea persoanelor

Pe baza rezultatelor obținute la test se decide în care clasă, din mai multe existente, să fie repartizat fiecare subiect. În acest mod se procedează atunci când:

- se alege cel mai potrivit program de pregătire școlară sau profesională, în funcție de particularitățile psihice și de cunoștințele persoanelor;
- la sfârșitul unui program de instruire sau cu ocazia examinării pentru admiterea la o formă de învățământ, se decide dacă un candidat poate fi declarat "reușit" sau "nereușit";

- se identifică persoanele care au caracteristici speciale (de exemplu sunt capabile de performanțe superioare într-un anumit domeniu), folosind un test de selecție simplu, care se administrează rapid (screening test); pentru persoanele care au realizat scoruri mai mari decât un prag fixat se vor administra apoi alte teste, care realizează măsurători mai precise.

B. Diagnoză psihologică și alegerea tratamentului

Diagnoza urmărește să determine natura, cauza și intensitatea comportamentului anormal al persoanei și să identifice categoria de boală în care este cel mai potrivit să fie încadrat subiectul (diagnosticul). Pe baza informațiilor furnizate de diagnoză se instituie tratamentul pentru remedierea deficiențelor sau pentru îmbunătățirea performanțelor.

Adesea, testele de personalitate servesc la diagnoza tulburărilor emoționale.

C. Autocunoaștere

Comparând rezultatele obținute la testele psihologice cu cele ale altor persoane, subiectul se poate cunoaște mai bine.

D. Evaluarea eficienței unui program educațional sau social

O modalitate de evaluare a eficienței constă în compararea rezultatelor obținute la teste psihologice de subiecții care au fost supuși influenței programului, fie cu rezultatele unor persoane care nu au fost incluse în programul respectiv, fie cu scorurile aceluiași persoane, înainte de acțiunea programului.

E. Cercetări teoretice sau aplicative asupra comportamentului uman

Se folosesc diverse teste psihologice pentru a depista cauza unor comportamente sau condițiile - de natură psihică - în care un anumit tratament (de exemplu, o metodă de psihoterapie sau un program de instruire) are șanse de succes.

Construirea unui test psihologic impune efectuarea unei cercetări în care se folosesc și teste psihologice cunoscute. Un același test psihologic poate fi utilizat în scopuri diferite. De exemplu, un test care permite formularea unui diagnostic psihiatric poate servi și la autocunoaștere.

1.3. CONDIȚII DE UTILIZARE A TESTELOR PSIHOLOGICE

În mai multe țări au fost elaborate norme etice și standarde de utilizare a testelor psihologice și de practicare a examinărilor psihologice.

Standardele citate mai frecvent în literatura de specialitate și la care vom face referire în această lucrare sunt cele elaborate în S.U.A., în 1985, de reprezentanți a trei organizații profesionale: *The American Educational Research Association*, *The American Psychological Association (APA)* și *The National Council on Measurement in Education*. Intitulate *Standards for Educational and Psychological Testing*, ele descriu amănunțit condițiile care trebuie îndeplinite la construirea, evaluarea, administrarea și interpretarea testelor psihologice. Sunt vizate două aspecte, și anume:

- caracteristicile psihometrice ale testelor;
- comportamentul profesional al psihologilor, insistându-se asupra ideii că ei poartă responsabilitatea pentru apărarea bunăstării persoanelor testate.

Aceste Standarde au fost completate cu prevederi specifice unor domenii particulare. Așa sunt, de exemplu:

- *Specialty Guidelines* (APA, 1981), cu îndrumări pentru psihologii care lucrează în clinică, în organizații industriale, în școli sau în activitatea de consiliere;

- *Principles for the Validation and Use of Personnel Selection Procedures* (Society for Industrial and Organizational Psychology, 1987), privind condițiile de desfășurare a selecției profesionale;

- *Guidelines for Computer-Based Tests and Interpretations* (APA, 1986), referitoare la folosirea calculatorului în testarea psihologică.

În ceea ce privește principiile etice care trebuie să îi călăuzească pe psihologi în activitatea de evaluare, în completarea celor cuprinse în *Standards for Educational and Psychological Testing* vin cele enunțate în *Ethical Principles of Psychologists and Code of Conduct* (APA, 1992). Redăm câteva dintre ideile enunțate aici:

- evaluarea psihologică și psihodiagnosticul trebuie să se facă numai într-un cadru profesional, de către utilizatori de teste competenți și cu experiență în domeniu, cu ajutorul unor teste potrivite;
- pentru proiectarea și selectarea tehnicilor adecvate populațiilor particulare se vor folosi proceduri științifice;
- interpretarea rezultatelor și explicarea acestora se vor face cu mare atenție;
- se va acționa pentru păstrarea securității testelor.

În proiectul *Statutului psihologului din România*, propus de biroul Asociației Psihologilor din România, se precizează că profesia de psiholog "se

exercită numai de către psihologi" (art. 1). Această profesie constă în observarea, evaluarea, explicarea, predicția și modificarea comportamentului uman prin aplicarea principiilor, metodelor și procedurilor științifice în scopul prevenirii și tratamentului tulburărilor psihice și de comportament, precum și scopul dezvoltării relațiilor interpersonale, adaptării eficiente la viață și muncă, optimizării ergonomice a producției, creșterii eficienței personale și menținerii sănătății psihice și fizice (art. 2).

Vom prezenta în continuare câteva aspecte care sunt deosebit de importante atunci când se utilizează teste psihologice (Anastasi, 1976; Gregor, 1992; Murphy, Davidshofer, 1991).

A. Calificarea examinatorului

Testarea psihologică efectuată de nespecialiști este extrem de dăunătoare, atât pentru subiect, cât și pentru cel care a solicitat examinarea (organizatorul unui program de instruire care trebuie să trieze candidații, compania care dorește să angajeze personal etc.), întrucât există riscuri mari să se ajungă la concluzii false. Din acest motiv, *La Commission Internationale de Reglementation de la Production et de la Vente des Tests*, la Congresul Asociației Internaționale de Psihologie Aplicată, în 1971, a decis să reglementeze distribuția testelor psihologice astfel:

- **autorilor** li se cere să limiteze distribuția testelor la persoanele care, prin formația lor și prin respectul manifestat față de metodologia și deontologia profesiei lor, sunt capabile să le folosească;
- **editorilor** li se cere să controleze distribuția testelor, vânzându-le direct numai utilizatorilor calificați profesional;
- **utilizatorilor** li se cere să aplice testele numai conform indicațiilor date de autorii acestora. Ei trebuie să respecte drepturile de autor, să nu reproducă, traducă, adapteze sau modifice un test fără permisiunea autorului și a editorului. Persoanele a căror formație profesională este incompletă nu pot cumpăra și utiliza teste decât prin intermediul și sub supravegherea unei persoane care are calificarea necesară.

Aceste hotărâri sunt respectate de toate marile firme care produc, distribuie sau utilizează teste psihologice. De exemplu, *Etablissement d'Applications Psychotechniques (EAP)*, o instituție franceză care echipează cu teste psihologice cele mai mari centre de examinare psihologică din Franța (SNCF, Citroen etc.) și din lume (România, Polonia, Belgia, Olanda etc.) obligă persoanele care doresc să cumpere teste psihologice să prezinte o fotocopie a diplomei de licență în psihologie.

Numeroase state au acceptat ca dreptul de profesare a psihologiei să fie acordat numai pe baza unor standarde de competență, cum sunt titlul de doctor

În psihologie, diploma de master în psihologie sau diploma de absolvire a unui curs de specializare într-un domeniu al psihologiei. Diploma de absolvire a facultății de psihologie este adesea insuficientă.

De exemplu, *Guidelines for Computer-Based Tests and Interpretations* (APA, 1986) apreciază că au dreptul de utilizare a testelor psihologice doar "profesioniștii calificați" care, în plus, îndeplinesc și următoarele condiții (după Eyde, Kowal, 1987):

- au cunoștințe asupra măsurării psihologice;
- cunosc istoricul testelor sau al chestionarelor pe care le utilizează;
- au experiență în utilizarea testelor, astfel încât pot aprecia când există diferențe de sex, vârstă sau cultură între rezultatele la un test;
- au cunoștințe în domeniul în care aplică testele.

De exemplu, în cazul chestionarelor de personalitate, utilizatorul trebuie să aibă cunoștințe de psihopatologie și de teoria personalității.

Cataloagele care prezintă teste psihologice specifică și nivelul de competență necesar cumpărării și utilizării fiecărui test. În multe țări, studenții pot achiziționa teste psihologice, doar dacă le prezintă distribuitorilor de teste semnătura persoanei responsabile de activitatea lor.

B. Calitatea instrumentelor și a procedurilor de măsurare

Înainte de a utiliza un test, psihologul are obligația de a citi manualul testului, pentru a afla dacă:

- testul este adecvat scopului urmărit (de exemplu, dacă măsoară trăsătura care prezintă interes);
- testul este potrivit caracteristicilor subiectului /subiecților (sex, vârstă, nivel de studii etc.);
- calitățile testului (validitate, fidelitate, dificultate etc.) sunt acceptabile.

Este important de reținut că felul în care acționează un test depinde nu numai de calitățile sale, constatate în diverse experimente, ci și de scopul pentru care este folosit și de celelalte procedee de examinare psihologică utilizate împreună cu el. De exemplu, s-a constatat că, în cazul selecției profesionale, subiecții au mai multă încredere în informațiile furnizate de testele de aptitudini decât în cele obținute din chestionarele de personalitate, ceea ce influențează negativ sinceritatea răspunsurilor la itemii chestionarelor. S-a observat, de asemenea, că asocierea chestionarelor de personalitate cu teste de aptitudini și cu un interviu creează o atitudine mai favorabilă față de examenul psihologic decât atunci când chestionarele de personalitate sunt însoțite doar de interviu (Fletcher, 1997).

Testarea psihologică necesită pregătiri pentru asigurarea calității bune a materialelor utilizate și a condițiilor de examinare.

Un test redactat neglijent, cu greșeli de ortografie, pătat sau cu figuri desenate neglijent le va face o impresie proastă subiecților, scăzându-le interesul pentru completarea sa ori, uneori, derutându-i. Performanțele la test vor fi, astfel, afectate.

Examenele psihologice improvizate în săli lipsite de mese, gălăgioase sau cu o luminozitate improprie vor conduce, de asemenea, la distorsionarea rezultatelor.

Este indicat ca examenul psihologic să fie anunțat din timp și să se specifice cum trebuie să se pregătească subiecții: să nu se prezinte la examen obosiți, să nu consume alcool înaintea examinării, dacă folosesc ochelari, să îi aibă asupra lor etc.

C. Obținerea consimțământului subiectului pentru examinarea psihologică

Psihologul are obligația să îi explice subiectului motivul examenului psihologic și să obțină acordul acestuia de a se supune la test. Între persoana testată și examinator se va încheia o convenție scrisă, semnată de ambele persoane, prin care subiectul declară că a primit explicații ample despre scopul examinării, despre testele care vor fi utilizate și despre modul în care vor fi folosite rezultatele, își dă consimțământul să fie examinat psihologic și indică persoanele cărora le pot fi comunicate rezultatele examinării.

Dar, obținerea consimțământului unei persoane pentru a participa la o examinare psihologică poate denatura rezultatele testelor, întrucât, cunoscând ceea ce se intenționează să se măsoare, subiectul poate răspunde nesincer, din dorința de a crea o anumită imagine despre sine. Psihologul este pus uneori în situația de a opta între a obține date corecte despre indivizii testați și a respecta dreptul celor examinați de a fi informați despre test. În general, atunci când intruziunea testului în viața particulară este minoră și se asigură confidențialitatea răspunsurilor nu li se mai dau subiecților informații detaliate despre test.

D. Confidențialitatea rezultatelor

Cu excepția cazului când persoana examinată reprezintă un pericol pentru sine sau pentru alții, informația furnizată de testele psihologice este confidențială și nu va fi comunicată altora fără consimțământul subiectului. Chiar și atunci când subiectul este de acord ca rezultatele sale să fie cunoscute și de alții, rezultatele vor fi arătate doar medicului, avocatului sau psihologului care se ocupă de persoana respectivă.

Acceptând ideea că scorurile realizate la un test pot fi afectate de erori, uneori mari, care pot cauza interpretări neconforme cu realitatea, cu ocazia

comunicării rezultatelor trebuie să i se dea posibilitate subiectului să corecteze ceea ce el consideră că este o impresie nefavorabilă despre sine (Fletcher, 1997). Din acest motiv, în unele țări, cum este Olanda, codul etic pentru psihologi prevede ca rezultatele testării să fie comunicate mai întâi subiectului.

În mod obișnuit trebuie evitată afișarea rezultatelor examenului psihologic. De exemplu, în cazul selecției profesionale se recomandă să se afișeze doar numele persoanelor declarate admise, fără alte date care ar putea constitui o violare a intimității subiecților. Cei nereușiți nici nu trebuie amintiți. Acest lucru se impune mai ales atunci când sunt utilizate teste de inteligență sau de personalitate, întrucât denumirile acestora pot sugera interpretări care să îi lezeze pe subiecții cu scoruri foarte mici (sau foarte mari, în funcție de test).

E. Comunicarea rezultatelor testării

Niciodată rezultatele unui examen psihologic nu sunt comunicate sub forma lor brută, ci doar prelucrate și interpretate. Obișnuit, ele sunt prezentate în cadrul unui *raport psihologic*, întocmit în conformitate cu solicitarea beneficiarului. Acest raport va fi redactat clar, fără ambiguități. Se va evita utilizarea termenilor tehnici, dificili de înțeles de nespecialiști. De asemenea, nu se va face presupunerea că beneficiarul cunoaște testul și îi sunt suficiente doar informații sumare. De exemplu, unui nespecialist trebuie să i se explice ce înseamnă faptul că o persoană a obținut un coeficient de inteligență de 87 și ce implicații ar putea avea aceasta. Datele comunicate trebuie să reflecte obiectiv situația, evitându-se observațiile subiective și părerile personale.

O deficiență a multor rapoarte psihologice constă în faptul că ele descriu comportamente (de exemplu, "în timpul discuțiilor, subiectul gesticulează continuu"), fără a oferi comentarii asupra lor. Aceasta se întâmplă, de obicei, atunci când nici psihologul nu a putut interpreta faptele sau a considerat că explicațiile nu prezintă interes pentru cel care citește raportul. Dar, simpla enumerare a observațiilor făcute, fără comentarii, poate genera o mare varietate de interpretări și poate conduce la formarea unei imagini greșite asupra subiectului. Este preferabil ca în raport să fie incluse toate explicațiile pe care psihologul le consideră posibile, atrăgându-se atenția că ele nu sunt certe.

Uneori, rezultatele obținute la teste de o persoană sunt contradictorii. De exemplu, subiectul poate să apară la un test ca introvertit, iar la altul, ca sociabil. În asemenea situații este important ca psihologul să prezinte toate constatările făcute și să încerce să găsească explicații.

F. Efectul discriminatoriu al testului asupra grupurilor minoritare

Aici termenul de "grup minoritar" este legat de contextul în care se aplică testul și definește un grup, format în funcție de o anumită variabilă (sex,

statut social, nivel socio-cultural, apartenență etnică etc.), care conține mult mai puține persoane decât celelalte grupe. De exemplu, la selecțiile profesionale pentru postul de secretară, faptul că numărul bărbaților care se prezintă este, de obicei, mult mai scăzut decât cel al femeilor, face ca bărbații să constituie un grup minoritar.

Problema discriminării produse de utilizarea unui test se pune mai ales atunci când testul respectiv este folosit în cadrul selecției profesionale și se utilizează un singur prag de admisibilitate (cutoff score), adică o singură valoare cu care se compară scorurile subiecților pentru a se decide dacă sunt admiși sau respinși. Dacă există grupuri minoritare care obțin scoruri mai mici decât alte grupuri, persoanele care fac parte din ele vor avea șanse mai mici de a fi selectate. Ele se vor considera victime ale discriminării. Realitatea nu este totdeauna aceasta. Să analizăm două situații în care se prezintă la selecție două grupuri, unul cu scoruri mari la test și celălalt cu scoruri mici.

(1) Să presupunem mai întâi că testul este construit astfel încât rezolvarea sa nu este influențată de elemente care nu au nici o legătură cu variabila prezisă de test, numită criteriu. De exemplu, testul este utilizat la selecția profesională pentru un post de casier; el se referă la deprinderea de a efectua rapid și corect calcule aritmetice și rezolvarea sa nu este dependentă de gradul de stăpânire a limbii. Criteriul utilizat fiind performanța în muncă, dacă linia de regresie a criteriului față de test este aceeași pentru ambele grupuri, atunci din faptul că un grup obține scoruri mai mici la test se deduce că el va avea rezultate mai slabe și în muncă, deci este normal ca persoanele din grupul respectiv să fie respinse la selecție. În acest caz testul nu are efect discriminatoriu.

(2) Se poate întâmpla, însă, ca cele două grupuri să fie la fel de potrivite pentru postul pentru care se face selecția (performanțele lor în muncă sunt similare), dar liniile de regresie ale criteriului față de scorurile testului să difere de la un grup la altul. Astfel de situații apar atunci când persoanele dintr-un grup obțin rezultate slabe la test dintr-un motiv care în activitatea viitoare nu are importanță sau poate fi compensat (de exemplu din cauza unui handicap de limbă ori cultură). Dacă se folosește același prag de admisibilitate pentru ambele grupuri, vor fi respinse unele persoane cu scoruri mici la test care ar fi putut obține performanțe în muncă mai bune decât unele persoane admise. În asemenea cazuri se poate vorbi despre efectul discriminatoriu al testului, întrucât sunt defavorizați la selecție cei din grupul cu scoruri mici la test. Pentru acest grup ar fi trebuit fixat un prag de admisibilitate mai scăzut.

CAPITOLUL 2

CÂTEVA NOȚIUNI CARE INTERVIN ÎN TEORIA TESTELOR PSIHOLOGICE

2.1. CONSTRUCTUL PSIHOLOGIC

Un construct este, în general, o idee "construită" de experți dintr-un domeniu particular al științei, pentru a rezuma un grup de fenomene sau de obiecte și pentru a fi utilizată într-un cadru științific (teoretic, metodologic sau aplicativ).

După Murphy și Davidshofer (1991), orice construct are două proprietăți:

- este o abstractizare a unor regularități din natură;
- nu este observabil direct, dar poate fi conectat cu entități sau evenimente concrete, observabile.

Constructele psihologice sunt evidențiate, direct sau indirect, de comportament sau în urma conducerii unor experimente dedicate măsurării lor. Un construct psihologic, arată J.F. Binning și G.V. Barrett (Silva, 1993), este "o etichetă aplicată unei grupări de comportamente care covariază". Pentru *Standards for Educational and Psychological Tests* din 1985, constructul este "o noțiune teoretică derivată din cercetare și din alte experiențe, care a fost construită pentru a explica patternuri de comportament observabile".

Constructele create pentru interpretarea rezultatelor testelor psihologice sunt identificate, de obicei, cu câte o clasă de răspunsuri (L. J. Cronbach, după Silva, 1993).

Unii autori consideră constructele ca fiind multidimensionale, formate din "componente" independente unele de altele, numite uneori "fațete", și recomandă utilizarea analizei factoriale ca metodă pentru studierea acestora. De exemplu, în modelul elaborat de Guilford, constructul "creativitate" are drept componente: fluiditatea, flexibilitatea, originalitatea, elaborarea, senzitivitatea față de probleme și redefinirea (Landau, 1979). Prezența componentelor ridică probleme în studiul relației dintre construct, privit ca variabilă independentă, și o variabilă dependentă prezisă de construct. O posibilitate de rezolvare este cotarea separată a componentelor, cumulara scorurilor și cercetarea relației dintre scorul total obținut și valorile variabilei dependente. Dar, uneori, componentele ar putea să prezică mai bine variabila dependentă decât constructul general. În asemenea situații se studiază separat relația dintre fiecare componentă și variabila dependentă (Ozer, Reise, 1994).

Fiecare construct psihologic este fundamentat pe o teorie, care permite descrierea și predicția comportamentelor legate de el, în situații specifice.

Pentru elaborarea unui test care să îl măsoare, constructului i se asociază o variabilă cantitativă despre care se presupune că se află în relație crescătoare cu scorurile testului. Se formulează apoi o serie de afirmații referitoare la modul în care se comportă persoanele care au valori mari ale acestei variabile, comparativ cu cele care au valori scăzute. De exemplu, vrând să dezvolte un test de evaluare a motivației indivizilor pentru propria realizare, McClelland și colaboratorii săi au formulat următoarele descrieri de comportamente (Thorndike, Haagen, 1961):

- Persoanele care au o motivație puternică pentru realizare au rezultate școlare bune.
- Motivația pentru realizare este mai mare dacă subiecților li se administrează în prealabil teste despre care li se spune că măsoară inteligența.
- Cei care au o motivație puternică pentru realizare completează mai mulți itemi într-un test de viteză a cărui funcție le-a fost explicată.
- Motivația pentru realizare este mai mare la copiii a căror familie le-a încurajat de timpuriu independența.

Fiecare dintre aceste descrieri este bazată pe un fel de "teorie a comportamentului uman". De exemplu, realizarea școlară este văzută ca o combinație între aptitudinea școlară și efort. Dintre persoanele cu același nivel aptitudinal, cele care au o motivație mai mare pentru realizare depun un efort mai mare și obțin rezultate școlare mai bune (Thorndike, Haagen, 1961).

Comportamentele prin care autorul testului consideră că se manifestă constructul pot fi influențate de diverse variabile, precum vârsta, sexul, educația primită etc. Pentru unele dintre acestea legătura cu constructul poate corespunde unei situații reale, observată în populația căreia i se adresează testul și acceptată de teoria care stă la baza constructului. Pentru altele, însă, ea denotă că aceste comportamente nu sunt caracteristice constructului. De asemenea, comportamentele prin care este descris constructul măsurat de test se pot asocia cu comportamente prin care se manifestă alte constructe psihologice, evaluate prin teste existente.

În consecință, elaborarea unui test care să măsoare un anumit construct impune descrierea constructului în termeni comportamentali concreți. Operația se numește *explicarea constructului* și constă din următorii trei pași (Murphy, Davidshofer, 1991):

- (1) Se identifică acele comportamente care au legătură cu constructul măsurat de test.

(2) Se identifică alte constructe și pentru fiecare se decide dacă are sau nu legătură cu constructul măsurat de test.

(3) Pentru fiecare construct se alcătuiește o listă de comportamente prin care acesta se manifestă. Pentru fiecare comportament, pe baza relațiilor dintre constructe, se decide dacă are sau nu legătură cu constructul măsurat de test.

Rezultă astfel un sistem de relații care înglobează un set de constructe și comportamente, denumit de L. J. Cronbach, *rețea nomologică* ("nomologică" în sensul că ea are la bază declarații exprimate sub formă de legi). Unii autori au folosit termenul de "rețea stohastică", susținând că relațiile dintre componentele rețelei sunt mai degrabă probabilistice decât de tip lege.

O descriere amănunțită a constructului trebuie să conțină și relațiile acestuia cu cât mai multe variabile fiziologice, psihologice etc. De exemplu, dacă se urmărește realizarea unui test pentru evaluarea inteligenței, se pot avea în vedere următoarele relații ale constructului "inteligentă" cu alte variabile (Gronlund, 1967; Kline, 1993):

- Scorurile testului cresc cu vârsta (se presupune că inteligența crește cu vârsta aproximativ până la 16 ani).
- Scorurile testului prezic succesul școlar.
- Scorurile testului concordă cu aprecierile făcute de profesori despre inteligența elevilor.
- Scorurile testului corelează pozitiv cu scorurile unor teste așa-numite "de inteligență".
- Dacă se efectuează analiza factorială asupra acestui test și a altor teste care măsoară aptitudini, se va constata că testul de inteligență are o saturație mare în primul factor general.
- Scorurile testului discriminează între un grup de copii handicapați mintal și unul format din copii capabili de performanțe superioare.
- Scorurile testului sunt influențate doar puțin de procesul instrucțional urmat de subiect.
- Scorurile testului discriminează între diverse grupe ocupaționale. Unele profesii, precum cea de matematician sau de funcționar superior, necesită o inteligență suuperioară, în timp ce altele pot fi exercitate și de persoane cu o inteligență redusă.

Fiecare dintre afirmațiile formulate în legătură cu constructul pe care testul trebuie să îl măsoare urmează să fie verificată experimental. Dacă unele dintre ele nu se confirmă, trebuie revăzută teoria care stă la baza constructului, pentru a sesiza eventualele omisiuni sau neconcordanțe.

2.2. DOMENIUL DE CONȚINUT AL UNUI TEST

Prin **domeniu de conținut al unui test** se înțelege "mulțimea tuturor comportamentelor care pot fi utilizate pentru a măsura atributul specific sau caracteristica la care se referă testul" (Murphy, Davidshofer, 1991).

Definirea domeniului de conținut este asemănătoare celei de definire a unui construct. Autorul testului, pe baza ideii pe care și-a format-o despre ceea ce vrea să măsoare, reține acele comportamente care presupune că ar fi manifestări ale constructului, respectiv, în cazul domeniului de conținut, expresii ale performanței sau ale cunoștințelor subiecților.

Deosebirea dintre un construct și un domeniu de conținut constă numai în gradul de abstractizare implicat și adesea este dificil de făcut distincție între ele (Guion, 1986).

Descrierea domeniului de conținut trebuie în primul rând să precizeze limitele acestuia. Apoi, pentru fiecare componentă se apreciază importanța. În cazul testelor destinate să măsoare rezultatele unui program de instruire (cunoștințe sau deprinderi), descrierea domeniului de conținut al testului se bazează pe analiza conținutului și a obiectivelor instrucționale ale cursului urmat de persoanele care vor fi testate. Se încearcă să se includă în domeniul de conținut aspectele pe care mai mulți experți le consideră importante.

Exemplu.

Un test de cunoștințe de psihometrie, administrat studenților din anul III de la facultatea de psihologie, după predarea cursului de Psihodiagnostic, poate avea ca domeniu de conținut "cunoașterea noțiunilor de bază ale teoriei testelor, conform materiei predate la curs".

Descrierea domeniului de conținut cuprinde:

- limitele domeniului: se au în vedere cunoștințele cuprinse în bibliografia indicată la curs (se anexează lista bibliografică);
- componentele domeniului:
 - probleme: analiza de itemi, validitatea testelor, fidelitatea testelor;
 - aspecte tratate: definițiile noțiunilor, metodele de cercetare, condițiile de utilizare a metodelor de cercetare și a formulelor de calcul;
- importanța relativă a componentelor domeniului: este indicată în tabelul 2.2.1. Procentele se referă la timpul alocat fiecărei probleme în cadrul cursurilor predate studenților.

Tabelul 2.2.1
Importanța componentelor domeniului de conținut
pentru un test de cunoștințe de psihometrie

Problema	Aspectul tratat		
	Definiții	Metode	Condiții de utilizare
Analiza de itemi	12%	14%	8%
Validitatea testelor	11%	17%	5%
Fidelitatea testelor	6%	19%	8%

Există, însă, multe domenii de conținut care nu pot fi descrise detaliat după modelul prezentat. De exemplu, domeniul de conținut al unui test de aptitudini cognitive poate fi format din itemi extrași din teste cunoscute, care se consideră că evaluează aptitudinea respectivă. Într-o asemenea situație este dificil de precizat care sunt limitele domeniului de conținut.

2.3. GRUPELE CONTRASTANTE

2.3.1. Definiția grupelor contrastante

În multe cercetări efectuate în psihologie și pedagogie se utilizează așa-numitele *grupe contrastante* sau *grupe extreme*. Este vorba despre două loturi de subiecți care, în privința uneia sau a mai multor variabile fixate de cercetător, au valori diferite: foarte mici într-o grupă, foarte mari în cealaltă.

Atunci când variabilele sunt măsurate pe scală de interval sau de raport, se poate da o definiție matematică a grupelor contrastante. În acest scop, se realizează mai întâi o ordonare a variabilelor, după un anumit criteriu (de exemplu, se așază în ordinea în care au fost înregistrate valorile lor). Apoi, pentru fiecare persoană, se transformă valorile variabilelor în cote standard z , utilizând mediile și abaterile standard ale acestor variabile în mulțimea tuturor subiecților (mulțimea din care se formează cele două grupe extreme). Fiecărei persoane i se pune în corespondență un vector, ale cărui componente sunt egale cu cotele standard ale variabilelor, în ordinea stabilită.

DEFINIȚIA 2.3.1.1. Prin centru de greutate al unei mulțimi de vectori

$$A = \{a_j = (a_{j1}, a_{j2}, \dots, a_{jn}) \mid j=1, 2, \dots, N\}$$

se înțelege vectorul

$$m = (m_1, m_2, \dots, m_n)$$

are cărui componente m_i sunt egale cu mediile aritmetice ale valorilor componentelor corespunzătoare, în mulțimea dată:

$$m_i = (a_{i1} + a_{i2} + \dots + a_{iN}) / N$$

$$i=1, 2, \dots, n.$$

DEFINIȚIA 2.3.1.2. (Albu, Pitariu, 1995c) Două grupe de subiecți se numesc **contrastante**, la un prag de probabilitate fixat, p , dacă sunt îndeplinite următoarele două condiții:

- Centrul de greutate al vectorilor corespunzători persoanelor unei grupe are toate componentele negative, semnificativ diferite de zero la pragul p , iar centrul de greutate al vectorilor corespunzători subiecților din cealaltă grupă are toate componentele pozitive, semnificativ diferite de zero la pragul p .
- Componentele corespunzătoare ale celor două centre de greutate diferă semnificativ între ele, la pragul p fixat.

Observații

- Prima condiție din definiția 2.3.1.2 asigură faptul că în grupa "bună" mediile valorilor variabilelor sunt superioare mediilor întregului lot, iar în grupa "slabă", ele sunt inferioare mediilor întregului eșantion.
- Cea de a doua condiție impune ca pentru fiecare variabilă mediile celor două grupe să fie semnificativ diferite între ele.

Fie (C_1, C_2) și (C_3, C_4) două perechi de grupe contrastante în sensul definiției 2.3.1.2, grupa C_1 având centrul de greutate

$$c_i = (x_{i1}, x_{i2}, \dots, x_{in}), \quad i=1, 2, 3, 4.$$

C_1 și C_3 sunt grupele "slabe", iar C_2 și C_4 sunt grupele "bune".

DEFINIȚIA 2.3.1.3. Se spune că elementele perechii (C_1, C_2) contrastează mai puternic decât elementele perechii (C_3, C_4) dacă are loc relația:

$$\min \{|x_{1j} - x_{2j}| : j=1, 2, \dots, n\} > \min \{|x_{3j} - x_{4j}| : j=1, 2, \dots, n\} \quad (2.3.1.1)$$

Se spune că elementele celor două perechi contrastează la fel de puternic dacă are loc relația:

$$\min \{|x_{1j} - x_{2j}| : j=1, 2, \dots, n\} = \min \{|x_{3j} - x_{4j}| : j=1, 2, \dots, n\} \quad (2.3.1.2)$$

3.2.2. Câteva teorii ale personalității

Deosebirea esențială dintre diversele teorii referitoare la personalitate constă în ponderea pe care o acordă rolului jucat de factorii interni (legați de *persoană*) și de cei externi (legați de *situație*) în declanșarea, menținerea și modificarea comportamentului. Căutând să răspundă la întrebarea "de ce variază comportamentul intrapersonal și cel interpersonal?", adepții acestor teorii au efectuat cercetări prin care au încercat să determine "când apar variații ale comportamentului". Fiecare teorie și-a constituit o metodologie proprie de cercetare, cu un mod specific de culegere și prelucrare a datelor.

Multă vreme, până prin anii 1960-1970, teoriile și cercetările referitoare la personalitate au fost dominate de **modelul trăsăturilor** (trait model). Conform acestuia, comportamentul (C) este determinat de predispoziții latente stabile ale persoanei (P), numite trăsături. Adică:

$$C = f(P).$$

Modelul trăsăturilor oferă posibilitatea de a concentra într-un set limitat de trăsături caracteristici afective, atitudinale și comportamentale ale indivizilor. Dar, fiind o teorie descriptivă, el nu identifică mecanismele care stau la baza trăsăturilor. În acest model, "un psihograf al unei trăsături este ca un buletin meteorologic, bun pentru a vă spune dacă trebuie să purtați umbrela, dar sărac în furnizarea de explicații de ce ar putea să plouă" (Ozer, Reise, 1994).

În decursul timpului au fost elaborate modele cu o singură trăsătură (single-trait models) și modele cu mai multe trăsături (multi-traits models).

Pentru modelele din prima categorie, trăsătura este un aspect specific al personalității, care îndeplinește următoarele condiții:

- este prezent la toate persoanele, dar în grade diferite;
- este invariabil intraindividual;
- joacă un rol important în determinarea comportamentului;
- poate fi măsurat cu precizie.

Aceste modele utilizează instrumente (scale, teste, chestionare) prin care evaluează câte o singură trăsătură. În lucrările referitoare la personalitate, despre o asemenea trăsătură se spune că este "de nivel inferior" sau "specifică". Specificitatea se referă la contextul în care este definită trăsătura, respectiv la ~~varietatea situațiilor în care trebuie observate comportamentele care servesc~~ la definiția trăsăturii respective. Cu cât numărul acestor situații este mai mic și ele sunt mai asemănătoare între ele, cu atât trăsătura are un grad de specificitate mai mare. Spre exemplu, se vorbește despre trăsătura "anxietate", dar și despre trăsăturile "anxietate față de calculator" și "anxietate față de teste". În primul caz trăsătura are o generalitate mai mare, o specificitate mai mică. Precizia cu care poate fi prezis un anumit comportament într-o situație particulară (cum ar fi

lucrul la calculator sau examenul psihologic) este mai mică pentru trăsăturile generale decât pentru trăsăturile specifice situației.

Unii autori au definit trăsături "legate de sarcină" (task-related personality traits). Este vorba despre trăsături, mai mult sau mai puțin generale, care pot servi la predicția succesului într-un anumit domeniu de activitate. Așa sunt, de exemplu, trăsăturile educaționale studiate de B. de Raad (1996).

Modelele cu mai multe trăsături operează la un nivel de abstractizare mai înalt decât modelele cu o singură trăsătură. Ele presupun că există un număr redus de trăsături fundamentale, globale, numite uneori "factori de nivel înalt" sau "superfactori", pe care încearcă să le identifice prelucrând prin analiză factorială rezultatele unor chestionare, observații și teste obiective referitoare la trăsături specifice. Factorii rezultați sunt considerați ca surse ale trăsăturilor specifice (Boekaerts, 1996). Cele mai importante modele cu mai multe trăsături sunt: modelul lui Eysenck bazat pe neuroticism, extraversiune și psihoticism, teoria tipurilor de personalitate și modelul Big Five.

În mod firesc, apare întrebarea: care trăsături este mai important să fie evaluate de către psihologi, cele specifice sau cele globale? Cercetătorii susțin că alegerea trebuie dictată de scopul examinării psihologice: pentru "explicare", sunt mai utile trăsăturile globale, în timp ce pentru "predicție", cele specifice (Lay, 1997).

În momentul definirii ei, fiecare trăsătură este pusă în legătură cu un grup de comportamente, care servesc ca indicatori ai acesteia. De exemplu, J.P. Guilford definește empatia ca "o abilitate de a prezice, de a recunoaște dispozițiile psihologice ale unei alte persoane, percepțiile sale, gândurile, sentimentele și atitudinile, de asemenea, ca trăsături proprii" (Marcus, 1997). Pentru aprecierea gradului în care trăsătura este prezentă la fiecare persoană, se utilizează un model aditiv, care constă din următorii pași:

(1) Se evaluează unul sau mai multe din următoarele criterii (G. W. Allport, după Corsini, 1994):

- frecvența cu care apar comportamentele;
- varietatea situațiilor în care apar;
- intensitatea cu care se manifestă;
- durata lor.

(2) Se cumulează măsurile obținute.

(3) Pe baza valorii rezultate, se fac inferențe asupra măsurii trăsăturii. În acest scop se aplică un algoritm specific instrumentului utilizat la evaluarea criteriilor sau se compară valoarea obținută pentru o persoană cu cele ale altor indivizi.

Testele și chestionarele folosite cel mai des pentru colectarea datelor, atât în cercetare cât și în psihologia aplicată (de exemplu, în clasificare, selecție,

consultanță) au la bază modelul trăsăturilor. De asemenea, multe metode matematice utilizate frecvent de psihologi pornesc de la ipoteza existenței unor variabile latente (corespunzătoare trăsăturilor) care sunt stabile în timp și se manifestă prin intermediul variabilelor observate. Așa sunt: analiza factorială, teoria clasică a testelor și teoria răspunsului la itemi.

Analiza factorială reduce o mulțime de variabile observate (scoruri de itemi, scale sau teste, rezultatele unor observații etc.), corelate liniar între ele, la un număr mai mic de variabile necorelate (factorii extrași), cu ajutorul cărora se poate realiza o descriere obiectivă a persoanelor. Un factor este o combinație liniară ponderată a variabilelor prelucrate prin analiză factorială. Ponderile rezultă din calcule și sunt determinate de valorile coeficienților de corelație liniară dintre variabilele măsurate. Factorii extrași sunt concepte statistice. Ei reprezintă variabile latente (ce nu pot fi observate direct), care au legătură de corelație liniară cu variabilele observate prelucrate prin analiza factorială. Ei nu corespund totdeauna unor caracteristici reale ale persoanelor sau unor trăsături. Semnificația fiecărui factor se deduce în urma analizei valorilor ponderilor variabilelor observate.

Teoria clasică a testelor folosește descompunerea scorului observat la test, scală sau item în *scor real* și în *eroare de măsură*. Scorul observat reprezintă rezultatul unei măsurători, în timp ce scorul real corespunde unei variabile latente. Dacă instrumentul utilizat în investigație servește la evaluarea unei trăsături, atunci scorul real este tocmai măsura trăsăturii respective. Teoria clasică a testelor utilizează în studiul relației dintre variabile modelul liniar. În consecință, gradul de acord dintre variabile este exprimat prin intermediul unui coeficient de corelație liniară.

În teoria răspunsului la itemi, spre deosebire de teoria clasică a testelor, relația dintre scorurile itemilor și valoarea variabilei latente măsurate de test este exprimată analitic printr-o funcție de regresie neliniară. Dacă itemul funcționează corect, funcția de regresie este monotonă (crescătoare sau descrescătoare).

Teoriile referitoare la personalitate au analizat variabilitatea comportamentelor sub două aspecte: al *consistenței* (constanța comportamentului în diverse situații existente la momente foarte apropiate între ele) și al *stabilității temporare* (constanța comportamentului în situații asemănătoare dar distanțate în timp). Cercetările efectuate au evidențiat grupe de comportamente asemănătoare în ceea ce privește variabilitatea intrapersonală și categorii de persoane pentru care unele comportamente se manifestă asemănător în privința variabilității.

Astfel, s-a observat că atât consistența cât și stabilitatea temporară sunt mai mari pentru comportamentele care sunt asociate cu inteligența sau cu

proces cognitive. Etichetele referitoare la personalitate, pe care le folosesc indivizii pentru a se descrie în scalele de autoevaluare, au o stabilitate temporară mare (Mischel, 1968). E. L. Kelly, administrând mai multe chestionare unor cupluri, de două ori, la un interval de 16-19 ani, a constatat că, spre deosebire de atitudinile față de căsătorie, care s-au schimbat mult și în diverse moduri, încrederea în sine, sociabilitatea, unele interese, unele valori economice și politice evaluate prin chestionare au avut o stabilitate temporară mare (Mischel, 1968).

Cercetările experimentale au condus la două observații importante:

- Persoanele se deosebesc între ele în ceea ce privește variabilitatea comportamentelor de la o ocazie (situație sau moment de observare) la alta, unele fiind mai schimbătoare, iar altele mai consecvente. Unii autori susțin că pentru fiecare aspect al personalității există trăsături generale care au rol moderator asupra variabilității intrapersonale. De exemplu, R. J. Larsen și colaboratorii săi au observat că variația ciclică săptămânală a dispozițiilor este mai evidentă la introvertiți decât la extravertiți (Penner, Shiffman ș.a., 1994).

- La un același individ, consistența comportamentală de la o situație la alta se manifestă asemănător pentru mai multe comportamente, nu doar pentru cele specifice unei anumite trăsături (Amelang, Borkenau, 1986). Pornind de la această observație, s-a definit o variabilă continuă, considerată de unii ca fiind o "trăsătură", care reprezintă consistența comportamentală generală. Persoanele cu valori mari ale acestei variabile sunt doar puțin dependente de situație.

S-a încercat să se determine cauza care face ca la o persoană comportamentele asociate cu unele trăsături să varieze mult de la o situație la alta, iar cele legate de alte trăsături, să fie constante. S-a observat că trăsăturile cărora le corespund comportamente puțin variabile sunt cele pe care indivizii le consideră "relevante" pentru ei (adică potrivite pentru a-i descrie) sau "observabile" (adică destul de evidențiate pentru a putea fi observate de alții) (Amelang, Borkenau, 1986). Probabil că, într-un profil al personalității care ar îngloba foarte multe trăsături, acestora le-ar corespunde punctele de extrem (de minim sau de maxim).

O posibilitate de evaluare a constanței (consistenței și stabilității temporare) unei trăsături pentru o persoană constă în analizarea similarității profilurilor corespunzătoare mai multor manifestări comportamentale, construite în situații diferite, și a similarității profilurilor corespunzătoare unor situații diverse, construite pentru diferite comportamente care au legătură cu trăsătura respectivă. Similaritatea perfectă este considerată o dovadă a constanței individuale (Schmitt, Borkenau, 1992).

Pe o poziție opusă modelului trăsăturilor se situează modelul situaționist (situationism model), ale cărui baze au apărut pe la începutul seco-

Dictionary (ediția din 1925), formată din aproape 18000 de descriptori, cuvinte care deosebesc comportamentul unei persoane de cel al alteia. Dintre acești termeni, 4504 se referă la trăsături. Se poate afirma că ei stau la originea modelului cu cinci factori.

În 1934, L. L. Thurstone a apelat la 1300 de persoane pentru a evalua, fiecare, un individ pe care îl cunoaște bine, folosind o listă compusă din 60 de adjective. Prelucrând prin analiză factorială rezultatele obținute, el a ajuns la concluzia că personalitatea poate fi explicată prin cinci factori independenți.

Pe de altă parte, R. B. Cattell, pornind de la lista cu cei 4504 termeni referitori la trăsături a lui Allport și Odbert, a aplicat analiza factorială și, prin rotație oblică, a obținut 12 factori, dintre care primii cinci apar cu saturații mari. Ei sunt greu de interpretat, dar Cattell susține că aceștia reprezintă cele mai importante dimensiuni ale personalității (John, 1990).

Adevărații descoperitori ai modelului Big Five sunt, însă, E. C. Tupes și R. E. Christal, care au analizat mai multe lucrări dedicate scalelor bipolare construite de Cattell și au găsit cinci factori de personalitate.

În anii următori, numeroase studii efectuate asupra unor chestionare de evaluare sau de autoevaluare și asupra unor liste de adjective ori de verbe, realizate pe eșantioane de subiecți de sexe, vârste și naționalități diferite și prin metode diferite de analiză au pus în evidență câte cinci factori, în majoritatea cazurilor asemănători între ei. Ei au fost denumiți de Lewis Goldberg, în 1981, *Big Five*, pentru că fiecare dintre ei acoperă un domeniu foarte larg și cuprinde un număr mare de caracteristici de personalitate, distincte între ele și specifice fiecărui factor.

Deși numele atribuit factorilor diferă de la un autor la altul, consensul în ceea ce privește conținutul psihologic al acestora este destul de mare. Prezentăm în continuare semnificația factorilor și câteva dintre denumirile sub care sunt cunoscuți (John, 1990; Zlate, 1994; Minulescu, 1996):

- Extraversia, sociabilitatea și ambiția, atracția interpersonală, afirmarea, puterea, izbucnirea (surgency) - arată capacitatea de orientare a personalității către exterior, modul de implicare în acțiune, sociabilitatea persoanei.

- Agreabilitatea, adaptabilitatea socială, plăcerea, nivelul socializării, amabilitatea, dragostea - se referă la calitățile emoționale ale persoanei și la comportamentele ei prosociale.

- Conștiinciozitatea, conformitatea, dependența, responsabilitatea, prudența, autocontrolul, interesul pentru muncă, puterea realizării - vizează modul concret, caracteristic al individului de a trata sarcinile, activitățile, problemele care apar în viața lui; cuprinde trăsături cum ar fi ordinea, disciplina, responsabilitatea socială.

- Stabilitatea emoțională, controlul emoțional, emoționalitatea, neuroticismul, adaptarea, anxietatea, afectul - arată caracteristicile emoționale ale persoanei (calm, mulțumit, emotiv) și diferite dificultăți emoționale (anxietate, depresie, iritabilitate).

- Cultura, intelectul, curiozitatea intelectului, inteligența, subtilitatea, deschiderea, independența, grija/atenția - se referă la funcțiile intelectuale (creativitate, inventivitate, deschidere la experiență).

Deosebirile existente între conținuturile factorilor și, în consecință, între denumirile atribuite lor, la diverși autori, provin din faptul că elementele introduse în prelucrare prin analiză factorială (cuvinte care exprimă trăsături, itemi, scale etc.) nu sunt indicatori puri ai factorilor (J. A. Johnson, F. Ostendorf, după Johnson, 1996). De obicei, fiecare element are o saturație foarte mare într-un factor ("factorul primar") și o saturație mare în alt factor ("factorul secundar"). Pentru a se determina ce măsoară factorii se grupează elementele care au același factor primar și se analizează conținutul acestora. Deosebirile dintre concluziile formulate de diverși autori sunt cauzate de faptul că factorii secundari la elementele grupate pentru un același factor primar diferă de la un studiu la altul. De exemplu, o scală alcătuită din itemi care se referă la conștiinciozitate, dar care au saturații mari pozitive în factorul extraversiune, va conduce spre o altă denumire a factorului "conștiinciozitate" decât o scală compusă din itemi care vizează conștiinciozitatea dar au saturații mari negative în factorul extraversiune. În primul caz denumirea dată factorului "conștiinciozitate" va indica o formă extravertită a conștiinciozității, iar în al doilea caz, o formă introvertită.

Într-un model al persoanei, prezentat în 1994, de R.R. McCrae și P.T. Costa, cei cinci factori din modelul Big Five sunt văzuți ca dispoziții psihice fundamentale, care au baze biologice, nu sunt observabile direct și oferă explicații pentru o categorie de fapte psihice numite "adaptări psihice". Acestea din urmă sunt caracteristici dobândite și includ obiceiuri, atitudini, deprinderi, valori și motive. La rândul lor, adaptările caracteristice furnizează explicații pentru comportamentele observabile (gânduri, sentimente, acțiuni, adică ceea ce evaluează de obicei persoanele din jur) (Minulescu, 1996).

Unele cercetări au arătat că fiecare dintre cei cinci factori este legat de bunăstarea psihică și de adaptarea din adolescență (Ozer, Reise, 1994).

Un model de evaluare a personalității acceptat de psihologi în ultimii ani constă dintr-o structură ierarhică: în vârf se găsesc cei cinci factori din modelul Big Five (numiți, uneori, *superfactori*), iar sub aceștia sunt plasate diverse fațete, care sunt măsurate prin instrumente ce se referă la aspecte înguste ale personalității (Goldberg, 1996).

- Se verifică dacă sunt îndeplinite condițiile de oprire a testării; în caz afirmativ se trece la pasul 3, iar în caz negativ se reia de la pasul 2. Testarea se termină de obicei când a fost administrat numărul de itemi fixat, când s-a epuizat timpul de testare, când intervalul de încredere calculat pentru variabila măsurată de test are lungimea mai mică decât o valoare fixată sau când acest interval nu mai include pragul de admisibilitate fixat.

(3) Se comunică subiectului nivelul determinat de test.

Algoritmul prezentat se referă la cazul când toți itemii testului măsoară o aceeași caracteristică. Dar, în teoria răspunsului la itemi au fost dezvoltat algoritmi care permit evaluarea unei persoane în privința mai multor variabile psihice.

La început, prin teste adaptative se evaluau doar aptitudini și cunoștințe. Acum se construiesc asemenea teste și pentru măsurarea unor variabile non-cognitive, cum sunt atitudinile.

4.2. CLASIFICĂRI ÎN FUNCȚIE DE MODUL DE ADMINISTRARE A TESTULUI

A. După numărul persoanelor care pot fi examinate simultan, testele pot fi:

- individuale;
- colective sau de grup.

Examenul psihologic individual este considerat neeconomic, subiecții fiind testați unul câte unul, ceea ce necesită mult timp. Testarea colectivă presupune examinarea simultană a mai multor subiecți.

Această clasificare nu este precisă, întrucât unele teste de grup pot fi aplicate și în mod individual (de exemplu, testul Raven). Situația inversă este imposibilă, mai ales în cazul testelor de performanță (Roșca, 1972).

Ambele moduri de examinare prezintă o serie de avantaje, dar și limite. Examenul psihologic individual îi permite psihologului să stabilească un contact mai bun cu subiectul, să ajungă mai repede la cooperare cu acesta. Totuși, decizia psihologului este mai încărcată de subiectivitate, iar examenul psihologic individual este mai puțin standardizat decât cel colectiv. În schimb, unele stări particulare ale subiecților, care pot denatura rezultatele testării, cum sunt oboseala, anxietatea, boala etc., sunt mai greu de detectat în condițiile examenului colectiv.

Testarea individuală i se potrivește mai bine psihologului clinician și celui de consiliere în anumite activități de orientare școlară și profesională.

Testarea colectivă este preferată în selecția profesională, când se operează cu grupuri mari de subiecți, și mai ales în examinările de tip screening.

B. După *timpul de execuție* a sarcinii impuse, testele de aptitudini și cele de cunoștințe pot fi:

- teste de viteză (speed tests);
- teste de randament (power tests):
 - cu limită de timp;
 - cu timp de lucru nelimitat.

Testele de viteză conțin itemi la care orice persoana din categoria celor cărora le este destinată proba poate să răspundă cu ușurință corect, dar care trebuie rezolvați rapid. De exemplu, trebuie barate numere sau figuri sau trebuie identificate figurile identice. Ele sunt administrate cu limită de timp, fixată astfel încât nici o persoană să nu poată rezolva integral testul. Performanța unui subiect este dată de numărul de itemi la care a răspuns. Aceste teste au la bază presupunerea că între viteza de lucru și gradul de dezvoltare a caracteristicii măsurate de test există o asociere pozitivă.

Testele de randament conțin itemi de dificultăți diferite, adesea așezați în ordinea crescătoare a dificultății. Timpul de rezolvare a unui asemenea test nu este limitat sau este suficient de generos pentru a permite parcurgerea integrală a testului de către toate persoanele, iar performanța unui subiect este indicată de numărul itemilor rezolvați corect.

Probele utilizate cel mai des sunt combinații ale celor două tipuri de teste. Ele conțin itemi de dificultăți diferite și sunt administrate cu limită de timp. Dacă timpul de lucru este atât de lung încât toate persoanele examinate au posibilitatea să rezolve toți itemii la care cunosc răspunsul, proba se comportă ca un test de randament. Dacă timpul de lucru este prea scurt și subiecții nu au posibilitatea să rezolve toți itemii la care știu răspunde, proba se comportă ca un test de viteză (Corsini, 1994).

Guilford (1965) apreciază că un test care a fost parcurs integral de cel puțin 75% dintre cei examinați poate fi considerat drept test de randament.

Atât testele de viteză, cât și cele de randament sunt astfel construite încât să nu permită realizarea unui scor maxim (egal cu numărul itemilor testului). Acest lucru este necesar în vederea obținerii unei imagini cât mai clare despre persoana testată, despre nivelul unei anumite aptitudini sau deprinderi a acesteia. Totodată, apare astfel posibilitatea unei ierarhizări mai precise a subiecților. În practică, distincția între testele de viteză și cele de randament este numai una de grad, cele mai multe teste depinzând de ambii factori în proporții variabile (Cascio, 1991).

mărimii coeficientului de corelație dintre scorurile testelor, se obișnuiește să se procedeze astfel (Aiken, 1994):

- Se împarte lotul de subiecți, aleator, în două părți egale.
- La prima testare, fiecare jumătate a lotului de subiecți primește un alt test.
- La a doua testare, se inversează testele administrate, astfel încât fiecare subiect să răspundă la ambele teste.
- Se calculează coeficientul de corelație liniară între scorurile celor două teste, fără a se lua în considerare momentul când au fost administrate testele.

Metoda formelor paralele se folosește și atunci când se dorește obținerea unei colecții de teste paralele, de bună calitate. Este, de exemplu, situația întâlnită în psihoterapie, când este necesară administrarea periodică, la intervale de timp relativ scurte, a unor teste de anxietate sau de depresie. Pentru ca subiecții să nu fie tentați să repete răspunsurile date la administrările anterioare ale testului se preferă utilizarea alternativă a unor forme de test paralele.

Un caz particular al metodei formelor paralele îl constituie metoda înjumătățirii (split-half). În literatura de specialitate, aceasta este inclusă uneori în cadrul analizei consistenței interne.

Metoda înjumătățirii constă în următoarele:

- Se administrează testul unui lot de persoane.
- Se împarte testul în două părți cât mai asemănătoare între ele.
- Se calculează coeficientul de corelație între scorurile observate la cele două jumătăți.

Dacă jumătățile testului sunt instrumente de măsură paralele, atunci valoarea calculată reprezintă coeficientul de fidelitate al fiecăreia dintre părți. Cu ajutorul lui, aplicând formula lui Spearman-Brown, se poate determina coeficientul de fidelitate al întregului test.

Metoda înjumătățirii rezolvă multe din problemele teoretice și practice inerente metodei test-retest și metodei formelor paralele. Mai întâi, nu este necesar să se administreze două teste diferite și nici să se administreze de două ori același test. În felul acesta nu există pericolul modificării scorurilor reale între administrări și nici cel al memorării și repetării răspunsurilor date la primul test.

Dar, dificultatea constă în construirea jumătăților care să fie măsurători paralele. Se recomandă (Cronbach, 1943; Pitariu, 1978; Traub, 1994) să se elaboreze perechi de itemi care să fie similari ca:

- formă (mod de prezentare a textului, a ilustrațiilor);
- conținut;
- tip de întrebare (cu răspuns liber, cu răspuns la alegere etc.);

- medie și abatere standard a scorurilor.

Dar, itemii trebuie să fie independenți, adică unul să nu sugereze răspunsul la celălalt.

Din fiecare pereche, un item va intra într-o jumătate de test, iar al doilea, în cealaltă jumătate.

Există multe posibilități de ordonare a itemilor în cadrul testului. Nu se recomandă plasarea celor două părți una după cealaltă, întrucât intervine oboseala persoanelor examinate, care va influența scorurile la cea de-a doua jumătate, iar dacă testul este administrat cu limită de timp, este foarte probabil ca mulți itemi din cea de-a doua parte să nu fie rezolvați datorită timpului de lucru insuficient și nu din cauza necunoașterii răspunsului corect. Pentru a evita asemenea situații, se poate adopta metoda par-impair: itemii unei jumătăți vor ocupa locurile cu număr de ordine par, iar ceilalți, locurile cu număr de ordine impar. Se obișnuiește să se plaseze itemii în ordinea crescătoare a dificultății lor (Lord, Novick, 1968).

Apare, însă, o problemă în cazul când testul se administrează cu limită de timp. Fiecare persoană care nu reușește să răspundă la ultimii 2-3 itemi va avea scorul zero la aceștia și, deci, scorurile la itemii pari și la cei impari vor apărea mai puternic corelate decât sunt în realitate. Din acest motiv, metoda înjumătățirii nu se va aplica atunci când administrarea testului se face cu limită de timp.

8.4.4. Coeficienți de fidelitate inter-evaluatori

În cazul testelor care nu au o cotare obiectivă, cum sunt testele proiective sau cele de creativitate, scorurile subiecților sunt influențate și de persoana care face evaluarea răspunsurilor. Chiar dacă evaluatorii îndeplinesc cerințele formulate în manualul testului, referitoare la cunoștințele și experiența necesare, este puțin probabil ca ei să atribuie scoruri identice fiecărui subiect.

Pentru a verifica în ce măsură scorurile testului sunt dependente de cel care a făcut cotarea, se calculează coeficientul de corelație liniară între scorurile atribuite acelorași subiecți de evaluatori diferiți. Un asemenea coeficient se numește coeficient de fidelitate inter-evaluatori.

Unii autori (Hammond, 1995) recomandă utilizarea, pentru aprecierea acordului dintre evaluatori, a coeficientului K al lui Cohen. Acesta ia valori între -1 și 1 și indică în ce măsură concordanța evaluărilor este diferită de cea datorată întâmplării (A se vedea Anexa I).

8.5. OBSERVAȚII ASUPRA METODELOR FOLOSITE PENTRU ESTIMAREA FIDELITĂȚII

Alegerea metodei aplicate pentru estimarea fidelității trebuie să țină seama de ceea ce măsoară testul, de scopul pentru care el va fi folosit și de condițiile concrete în care se poate face testarea (există sau nu posibilitatea administrării repetate a testului, există sau nu forme paralele ale testului etc.).

Atunci când testul măsoară o caracteristică durabilă și, deci, trebuie să fie puțin sensibil la situația de testare, sau când pe baza scorurilor testului urmează să se tragă anumite concluzii valabile pentru o perioadă de timp mai îndelungată, se va aplica metoda test-retest.

Dacă nu se poate asigura retestarea persoanelor după un interval de timp suficient de lung pentru ca subiecții să nu își mai poată aminti răspunsurile date la test, dar există o formă paralelă a testului, se va utiliza metoda formelor paralele.

În cazul particular când forma paralelă a testului a fost administrată împreună cu testul se va putea aplica metoda înjumătățirii.

Dacă se știe că testul se compune din itemi care măsoară aceeași caracteristică -mai mult sau mai puțin durabilă- pentru estimarea coeficientului de fidelitate se va folosi o formulă de calcul al unui coeficient de consistență internă.

În cazul testelor pentru care cotarea nu este obiectivă, evaluatorul devine o componentă a procesului de măsurare realizat cu ajutorul testului. Alături de o altă metodă de estimare a fidelității se va calcula și coeficientul de fidelitate inter-evaluatori.

În general valorile coeficienților de consistență internă sunt mai mari decât cele ale coeficienților de echivalență, care sunt, de obicei, mai mari decât coeficienții de stabilitate (Murphy, Davidshofer, 1987). Acest lucru se explică prin faptul că numărul factorilor care influențează erorile de măsură este mai mare în metoda test-retest și în metoda formelor paralele, decât în metoda analizei consistenței interne. De exemplu, acumularea de informații între test și retest va afecta numai valoarea coeficienților de stabilitate, nu și pe aceea a coeficienților de echivalență sau a coeficienților consistenței interne.

Dacă, însă, persoanele examinate țin minte răspunsurile date la test și le repetă la retest, se poate obține un coeficient de stabilitate mare, chiar dacă testul are consistență internă scăzută.

8.6. FACTORI CARE INTERVIN ÎN ESTIMAREA COEFICIENȚILOR DE FIDELITATE

Trebuie observat că experimentele efectuate pentru determinarea valorii unui coeficient de fidelitate nu cuprind, în general, întreaga populație de persoane căreia îi este destinat testul, ci doar un eșantion din această populație. Prin urmare, valoarea calculată este doar o aproximare (o estimare) a coeficientului de fidelitate. Luând un alt eșantion, se va obține o altă valoare.

Pentru ca estimarea coeficientului de fidelitate să fie cât mai precisă, este necesar să fie îndeplinite următoarele condiții (Traub, 1994):

A. Eșantionul să fie cât mai voluminos. Deși forma exactă a repartiției coeficienților de fidelitate calculați pentru diverse eșantioane de același volum, extrase aleator din populație, nu este cunoscută, se obișnuiește să se considere că abaterea standard a acestei repartiții (denumită eroare standard) este invers proporțională cu rădăcina pătrată din volumul eșantioanelor. Prin urmare, pentru a micșora de două ori eroarea standard este necesar să se mărească de patru ori volumul eșantionului.

B. Eșantionul de subiecți să fie reprezentativ pentru populația căreia îi este destinat testul. El trebuie să fie la fel de eterogen ca și această populație, întrucât gradul de omogenitate al eșantionului are influență asupra mărimii coeficientului de fidelitate.

C. Măsurătorile efectuate cu ajutorul testului să fie independente între ele, adică măsurătoarea realizată asupra unui subiect să nu fie influențată și nici să nu influențeze măsurătorile efectuate asupra altor persoane. Dacă administrarea testului se face în grup, acest lucru înseamnă că persoanele examinate vor fi supravegheate astfel încât nici una să nu poată copia.

Atunci când calculul coeficientului de fidelitate impune ca fiecare persoană să răspundă de două ori la același test sau să răspundă la două forme paralele ale testului foarte asemănătoare între ele, cerința de independență a măsurătorilor se extinde, în sensul că măsurătorile efectuate asupra aceleiași persoane trebuie să fie independente între ele. Aceasta înseamnă că nu li se va oferi subiecților posibilitatea de a revedea sau de a-și aminti răspunsurile date la prima administrare a testului.

Dacă se calculează un coeficient de consistență internă, se va urmări ca itemii testului să fie independenți între ei. Prin urmare, rezolvarea unuia nu va fi condiționată de rezolvarea altuia și nu vor exista itemi care să se refere la

același aspect, astfel încât subiecții să răspundă la ei în același mod (corect sau greșit).

D. Procedura de administrare a testului și a retestului sau a celor două forme paralele ale testului să fie identică și să fie aceeași cu cea care se va utiliza în practică. Astfel, de exemplu, dacă se prevede ca testul să fie administrat cu limită de timp, în 40 de minute, atunci acest interval de timp va fi respectat și la administrarea testului pentru estimarea fidelității.

8.7. INTERPRETAREA COEFICIENȚILOR DE FIDELITATE

Dintr-o mulțime de teste având toate celelalte caracteristici egale, este de preferat testul care are cea mai mare fidelitate. Dar, pe de o parte, testul cel mai fidel poate fi cel mai scump sau mai dificil de administrat. Pe de altă parte, caracteristicile testelor sunt rareori egale.

Este imposibil să se specifice un nivel minim care să fie impus coeficienților de fidelitate ai tuturor testelor. Dar, în practica obișnuită sunt utile următoarele recomandări:

A. Se vor cere niveluri înalte de fidelitate testelor folosite pentru a se lua decizii importante asupra persoanelor și celor care împart indivizii în mai multe categorii pe baza unor diferențe relativ mici între ei. În selecția profesională, de pildă, se vor folosi teste care au coeficienți de fidelitate mai mari decât 0.9 (J. C. Nunnally, după Hammond, 1995).

B. Se vor accepta niveluri scăzute de fidelitate atunci când testele se utilizează pentru a lua decizii preliminare, nu finale, și în cazul testelor care împart persoanele într-un număr mic de categorii, pe baza unor diferențe individuale mari.

C. Dacă un test servește la compararea grupelor de persoane, coeficientul său de fidelitate poate fi mai mic, chiar cuprins între 0.6 și 0.7. Dacă, însă testul se folosește pentru compararea persoanelor, una cu alta, coeficientul său de fidelitate trebuie să fie mai mare decât 0.85 (Aiken, 1994).

În general, fidelitatea chestionarelor de personalitate și a celor de interese este mai mică decât cea a testelor cognitive (de cunoștințe, de inteligență, de aptitudini speciale) (Aiken, 1994).

În practică se întâlnesc de obicei coeficienți de fidelitate cuprinși între 0.70 și 0.98 (Guilford, 1965). Testele cotate obiectiv care măsoară aptitudini cognitive pot furniza, în loturi eterogene de persoane, coeficienți de fidelitate mai mari decât 0.8. Testele cotate subiectiv care măsoară aptitudini, testele de cunoștințe și cele care evaluează performanțe tipice (de exemplu chestionare de personalitate) au, adesea, coeficienți de fidelitate mai mici decât 0.80 (Traub, 1994).

Testele standardizate de cunoștințe au, în general, o fidelitate mare spre moderată. În schimb, testele cu alegere multiplă utilizate la clasă pot avea o fidelitate mai scăzută. Coeficientul lor de fidelitate este apropiat de 0.75.

Validitatea fiind mai importantă decât fidelitatea, în practică se folosesc adesea teste care sunt valide pentru populația și situația în care sunt administrate, dar care nu au coeficienți de fidelitate foarte mari.

8.8. FIDELITATEA DIFERENȚEI SCORURILOR ÎNTRE TEST ȘI RETEST

De multe ori ne interesează modul în care variază scorurile observate la un test $m \in M$, administrat de două ori unui lot de persoane P (de exemplu, înainte și după un tratament). Procedul de lucru obișnuit constă în calculul mediei diferențelor individuale dintre scorul retestului și cel al testului și aplicarea unui test statistic pentru verificarea semnificației acestei medii. Se pune, însă, problema: cât de bine măsoară diferențele de scor schimbările reale petrecute în indivizi? Altfel spus, cât de fidelă este măsurarea diferențelor de scor?

Pentru a studia fidelitatea diferențelor de scor se definește, pentru fiecare persoană $p \in P$, variabila aleatoare D_{np} , ale cărei valori sunt toate diferențele de scor care ar putea fi observate la acea persoană:

$$D_{np} = X_{np,2} - X_{np,1} \quad (8.8.1)$$

$X_{np,1}$ și $X_{np,2}$ sunt variabilele aleatoare corespunzătoare scorurilor observate la test, respectiv la retest.

Făcând presupunerea obișnuită din teoria testelor, variabila D_{np} se compune dintr-o "valoare reală a diferenței scorurilor", T_{np} și o "eroare de măsură a diferenței scorurilor", E_{np} :

$$D_{np} = T_{np} + E_{np} \quad (8.8.2)$$

Variabila T_{np} este definită ca și medie a diferențelor observate:

$$T_{np} = M(D_{np}) \quad (8.8.3)$$

deci ea este egală cu diferența scorurilor reale ale persoanei p între test și retest.

Se notează:

D_{m*} = variabila aleatoare corespunzătoare diferenței scorurilor observate în întreaga populație P;

T_{m*} = variabila aleatoare care ia ca valori diferențele scorurilor reale între test și retest în populația P;

E_{m*} = variabila eroare de măsură pentru diferențele de scor în populația P.

Conform teoremei 6.1.1.1 în populația P are loc relația

$$D_{m*} = T_{m*} + E_{m*} \quad (8.8.4)$$

și se poate calcula fidelitatea diferențelor de scor pe baza formulei de definiție a coeficientului de fidelitate:

$$\rho^2(D_{m*}, T_{m*}) = \frac{\sigma^2(T_{m*})}{\sigma^2(D_{m*})} \quad (8.8.5)$$

Ținând seama de felul în care a fost definită variabila D_{m*} , se poate rescrie formula (8.8.5) astfel încât fidelitatea diferențelor de scor să fie exprimată cu ajutorul fidelității testului, ρ_1^2 , al fidelității retestului, ρ_2^2 , al dispersiilor în populația P a scorurilor observate la test și la retest, σ_1^2 și σ_2^2 , precum și a covarianței scorurilor observate la cele două administrări ale testului, σ_{12} (Traub, 1994):

$$\rho^2(D_{m*}, T_{m*}) = \frac{\sigma_1^2 \rho_1^2 + \sigma_2^2 \rho_2^2 - 2\sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}} \quad (8.8.6)$$

Se constată de aici că fidelitatea diferențelor de scor depinde crescător de fidelitatea testului și de cea a retestului, și depinde descrescător de mărimea covarianței scorurilor observate la test și la retest. Cu cât fidelitatea testului și cea a retestului sunt mai bune, cu atât fidelitatea diferențelor de scor este mai bună. Când scorurile observate la test și la retest sunt necorelate liniar ($\sigma_{12}=0$), fidelitatea diferențelor de scor este maximă.

CAPITOLUL 9 VALIDITATEA TESTELOR PSIHOLOGICE

9.1. DEFINIȚIA VALIDITĂȚII

Validitatea este unul dintre conceptele cele mai importante în activitatea de cercetare științifică. Ori de câte ori un cercetător își pune o întrebare de forma "Acest studiu răspunde într-adevăr problemei formulate?" sau "Acest instrument măsoară cu adevărat ceea ce vrem să măsurăm?" sau "Cercetările din acest laborator ne arată într-adevăr lumea reală?", el se interesează de fapt de validitatea studiului ori a instrumentului.

În esență, validitatea se referă la cât de bine realizează un studiu, o procedură sau un instrument de măsură ceea ce se presupune că face (Graziano, Raulin, 1993). În acest sens a fost înțeleasă validitatea, multă vreme, și în teoria testelor. Inițial, ea a fost definită și cercetată în legătură cu funcția îndeplinită de test. Astfel, de exemplu, Lafon (1973) definea validitatea unui test ca fiind calitatea acestuia "de a măsura exact ceea ce trebuie să măsoare", iar Lindeman (1978) scria că "validitatea unui test este dată de măsura în care acesta își îndeplinește funcția".

Observația că psihologii care utilizează un test nu sunt interesați, de fapt, de proprietățile acestuia, ci de atributele persoanelor testate, a condus spre o nouă accepțiune a termenului "validitate". Acum accentul nu mai cade pe integritatea testului, ci pe încrederea pe care o putem avea în deducțiile făcute pornind de la scoruri, luând în considerare întregul proces prin care s-au obținut aceste scoruri (Guion, 1986). Prin urmare, validitatea nu mai este privită ca o calitate a testului, ci ca o calitate a utilizării date rezultatelor testului. Sunt relevante, în acest sens, următoarele definiții:

- Validitatea este gradul în care o anumită inferență făcută plecând de la test are înțeles, este potrivită și utilă (*Standards for Educational and Psychological Tests*, 1985).

- Validitatea este o judecată evaluativă globală asupra gradului în care dovezi empirice și raționamente teoretice oferă sprijin ca, pe baza scorurilor testului, să se realizeze interpretări sau acțiuni potrivite (Messick, 1995).

- Validitatea unui test se referă la cât de potrivite sunt interpretările descriptive, explicative sau predictive care se dau scorurilor sale (Silva, 1993).

În această nouă accepțiune, validitatea unui test trebuie legată atât de test, de ceea ce măsoară itemii săi, cât și de contextul în care este utilizat testul și de persoanele care răspund la test. Este absolut necesar ca atunci când se

vorbește despre validitatea unui test să se precizeze funcția la care se referă și atributul măsurat sau variabila criteriu utilizată la compararea scorurilor. Nu este corect să se folosească expresia "validitatea testului", fără alte specificații. Nici un test nu este valid pentru toate scopurile sau în toate situațiile sau pentru toate eșantioanele de indivizi (*Standards for Educational and Psychological Tests*, 1974).

Exemplu.

Să considerăm un test de cunoștințe de matematică, ai cărui itemi conțin exerciții cu integrale.

Pentru a putea aprecia validitatea acestuia este necesar, în primul rând, să se analizeze testul, urmărind dacă măsurarea se face corect, adică dacă răspunsurile notate în grila testului sunt corecte și dacă itemii măsoară ceea ce se dorește. De exemplu, dacă se intenționează verificarea cunoașterii formulelor de calcul, itemii trebuie să conțină exerciții care se rezolvă prin simpla aplicare a acestor formule. În schimb, dacă se urmărește testarea deprinderii de a opera cu integrale, itemii vor conține exerciții care necesită efectuarea unor transformări pentru a se ajunge la formulele de calcul învățate.

În continuare se ia în considerare contextul în care se va aplica testul. El poate servi, de pildă, la evaluarea cunoștințelor de matematică ale absolvenților de liceu, la examenul de bacalaureat. Dar dacă se dorește să fie administrat la concursul de admitere la facultatea de psihologie, pentru a selecta persoanele care vor putea utiliza corect procedee statistice în prelucrarea datelor provenite din experimentele psihologice, este posibil ca predicțiile făcute pe baza scorurilor sale să nu fie corecte. Ar fi mai indicat să se folosească, în acest scop, un test de cunoștințe care să conțină exerciții cu probabilități.

Pe de altă parte, un test care conține exerciții cu integrale poate fi util pentru verificarea cunoștințelor de matematică ale candidaților la Universitatea Tehnică. Dar, dacă exercițiile sunt de dificultate medie, atunci la facultățile cu concurență foarte mare, unde se înscriu absolvenți de liceu foarte bine pregătiți, și la cele cu concurență foarte mică, unde se înscriu candidați cu o pregătire slabă, scorurile obținute de subiecți vor fi apropiate ca valoare și nu vor permite o diferențiere între persoane.

Dacă testul este administrat cu ajutorul calculatorului, atunci rezultatele sale pot fi influențate și de factori care nu au legătură cu cunoștințele de matematică (de pildă, de anxietatea față de calculator, care este legată de nivelul cunoștințelor de informatică, sau de deprinderea subiecților de a folosi tastatura). Prin urmare, este greșit să se aprecieze nivelul cunoștințelor de matematică pe baza scorurilor testului, dacă subiecții diferă între ei în privința experienței în lucrul cu calculatorul.

Validarea unui test este procesul prin care se investighează gradul de validitate a interpretării propuse de test (*Standards for Educational and Psychological Tests*, 1985). Orice validare trebuie să înceapă prin a răspunde la întrebările:

- Ce funcție îndeplinește testul?
- În ce scop se utilizează testul?
- În ce condiții se folosește testul?
- Ce caracteristici au persoanele cărora li se administrează testul?

Multă vreme, procedeele aplicate pentru validarea testelor au fost legate de răspunsul dat primei întrebări. Pentru fiecare dintre funcțiile testelor psihologice - măsurare, decizie și predicție - au fost formulate condițiile pe care trebuie să le îndeplinească un test valid și au fost propuse strategii de verificare a acestora. Fiecare set de condiții, cu strategiile corespunzătoare, a fost etichetat și a fost privit ca un "tip" de validitate sau ca o "față" a validității. În *Standards for Educational and Psychological Tests* din anul 1974 sunt prezentate trei tipuri de validitate:

- validitatea conceptuală (construct validity);
- validitatea de conținut (content validity);
- validitatea relativă la criteriu (criterion-oriented validity).

Validitatea conceptuală și validitatea de conținut sunt legate de funcția de măsurare a testelor. Un test este valid ca instrument de măsură dacă prin scorurile sale se realizează o măsurare corectă a atributului psihic pentru care a fost construit. Aprecierea corectitudinii măsurării impune identificarea tipului de scală pe care se face măsurarea și verificarea condițiilor din definiția acestuia.

Validitatea relativă la criteriu este legată de funcția de decizie și de cea de predicție. Un test este valid relativ la criteriu dacă pe baza scorurilor sale se pot lua decizii corecte sau se pot face prognoze corecte asupra persoanelor examinate. Strategia utilizată pentru a aprecia dacă un test este valid relativ la criteriu constă în compararea scorurilor sale cu valorile unei variabile criteriu. Se apreciază că testul este valid dacă el poate înlocui variabila criteriu, adică dacă asemănarea dintre scorurile sale și valorile variabilei criteriu este atât de mare încât, cunoscând scorul la test al unei persoane, se poate deduce valoarea variabilei criteriu pentru aceasta.

În practică nu este posibil totdeauna să se obțină valorile variabilei criteriu concomitent cu cele ale testului. Din acest motiv, strategiile folosite pentru aprecierea validității relative la criteriu au fost împărțite în două clase. Au apărut, astfel, două tipuri noi de validitate care, în unele lucrări de specialitate, înlocuiesc validitatea relativă la criteriu. Este vorba despre:

- **validitatea concurentă**, care necesită obținerea scorurilor la criteriu aproximativ în același timp cu cele ale testului;

- **validitatea predictivă**, care permite ca administrarea testului să precedă înregistrarea valorilor variabilei criteriu, între cele două acțiuni existând un interval de timp mai lung sau mai scurt, în funcție de situația concretă.

Standards for Educational and Psychological Tests din 1985 afirmă că nu se poate vorbi despre "tipuri" diferite de validitate: cele trei (sau patru) tipuri descrise anterior sunt doar simple etichete atribuite unor categorii de strategii de validare care se pot utiliza și asociate. De pildă, pentru un test de atenție concentrată se pot cerceta atât validitatea conceptuală (constructul măsurat fiind atenția concentrată), cât și validitatea relativă la criteriu (drept criteriu poate servi testul de atenție concentrată Toulouse-Pieron).

Ideea asocierii fiecărei categorii de strategii de validare cu o anumită funcție a testelor a fost abandonată. Se acceptă acum că validarea relativă la criteriu poate fi efectuată atât în cazul testelor care măsoară un construct, cât și pentru testele utilizate în predicție (Guion, 1986). În primul caz, criteriul este un construct ales datorită legăturilor pe care se presupune că le are (sau că nu le are) cu constructul măsurat de test. În cel de-al doilea caz, criteriul este o măsură a performanței sau a comportamentului în muncă. Nu se poate afirma că o strategie de validare este mai potrivită decât alta pentru o anumită utilizare a testelor sau pentru o categorie specifică de inferențe bazate pe scoruri.

C. L. Lawshe se pronunță pentru înlocuirea denumirii "tip de validitate" cu cea de "tip de analiză a validității"; prin "analiză a validității" el înțelege o procedură, un proces sau o strategie prin care se colectează sau se generează date în vederea determinării extinderii ori a forței inferențelor care pot fi făcute pe baza unei mulțimi de scoruri de test (Landy, 1986). Cu o singură modificare minoră, ce constă în înlocuirea expresiei "mulțime de scoruri de test" cu aceea de "mulțime de observații", definiția dată de Lawshe se transformă într-o definiție a testării ipotezelor. Se constată, astfel, că validarea testelor este o operație similară verificării ipotezelor științifice. În cazul validării testelor, ipotezele pot fi formulate astfel:

"Persoanele care au realizat un scor mare la testul X vor realiza o performanță bună în activitatea Y."

sau

"Persoanele care au obținut scoruri mari la testul X au un nivel înalt al constructului Y."

În continuare vom prezenta categoriile de strategii de validare. În locul denumirilor "tipurilor" de validitate întâlnite în majoritatea lucrărilor de specialitate din limba română:

- "validitate conceptuală" (concept validity) sau "validitate de construct" (construct validity)

și

- "validitate de conținut" (content validity)

vom utiliza denumirile

- "validitate relativă la construct"

și

- "validitate relativă la conținut",

care, alături de denumirea consacrată "validitate relativă la criteriu", sugerează că validitatea testului este cercetată sub unul dintre aspectele posibile (construct, conținut sau criteriu).

9.2. VALIDITATEA TESTULUI RELATIVĂ LA CONSTRUCTUL MĂSURAT DE ACESTA

Validitatea relativă la construct se aplică în cazul testelor care măsoară constructe psihologice, cum sunt chestionarele de personalitate. Ea face apel la diverse teorii psihologice pentru a verifica, în primul rând, dacă testul se referă cu adevărat la constructul pe care trebuie să îl măsoare, și apoi, dacă scorurile obținute de subiecți la test reflectă corect mărimea acestui construct la persoanele respective; altfel spus, dacă între scorurile testului și valorile constructului există o relație funcțională monotonă (de obicei această relație este presupusă liniară):

$$\text{scor la test} = f(\text{valoarea constructului})$$

Validitatea relativă la construct impune analizarea testului din mai multe puncte de vedere și prin metode diferite. Un algoritm posibil al validării relative la construct, valabil pentru toate măsurătorile psihologice și educaționale, inclusiv pentru evaluările performanței, constă din următorii pași (după Messick, 1995):

A. Mai întâi se cercetează cât de relevant și reprezentativ este conținutul testului pentru constructul care trebuie măsurat. O asemenea analiză impune descrierea amănunțită a *domeniului constructului* respectiv, adică prezentarea deprinderilor, a cunoștințelor, atitudinilor, motivelor, a altor atribute care au legătură cu constructul și specificarea relațiilor dintre ele. Un grup de experți apreciază apoi

dacă cele mai importante porțiuni ale domeniului constructului sunt acoperite de conținutul testului.

Este foarte utilă și cercetarea modului în care constructul este "reprezentat în test". Pentru aceasta se realizează un model al rezolvării testului, conținând mecanismele teoretice (proces, strategii, cunoștințe) despre care se presupune că ar avea influență asupra răspunsurilor date la test. Se analizează apoi dacă elementele modelului au legătură cu constructul pe care trebuie să îl măsoare testul și cu răspunsurile care se pot obține la test. De mare folos, în această analiză, poate fi rețeaua nomologică întocmită cu ocazia explicării constructului.

Dacă scorurile testului sunt dependente mai mult de alți factori decât de componentele constructului care trebuie măsurat, se vorbește despre o "subreprezentare a constructului în test". Acest lucru se poate întâmpla dacă elemente care nu au legătură cu constructul determină ca rezolvarea testului să fie foarte dificilă sau, din contră, foarte ușoară. De exemplu, dacă itemii unui test de cunoștințe de statistică pretind efectuarea unor calcule cu numere având multe zecimale, apare riscul de a obține răspunsuri eronate din cauza greșelilor de calcul și nu datorită absenței cunoștințelor de statistică.

B. O altă etapă a validării constă în analizarea consecvenței răspunsurilor la itemii testului. În acest scop se identifică procesele implicate în obținerea scorurilor la itemi și se urmărește dacă acestea acționează similar la toți itemii din întregul test sau dintr-o parte fixată a acestuia.

Mijloacele de verificare a consecvenței răspunsurilor sunt variate: se cercetează dacă timpii de răspuns la itemi sunt aproximativ constanți sau dacă scorurile itemilor corelează semnificativ între ele ori dacă mișcările ochilor subiecților în timpul conceperii răspunsurilor nu variază prea mult de la un item la altul.

C. Pornind de la ideea că scorul obținut la test este rezultatul acțiunii mai multor variabile, dintre care unele au legătură cu constructul măsurat, se urmărește dacă structura scorului la test reflectă structura domeniului constructului măsurat.

Pentru aceasta, se descompune scorul total al testului în scoruri parțiale (scoruri ale unor grupe de itemi care corelează între ei sau scoruri factoriale) și se presupune că fiecare asemenea scor reprezintă măsura unei variabile.

Adesea, în scopul identificării unor asemenea variabile se utilizează analiza factorială. Ea permite să se determine una sau mai multe variabile (numite factori) ale căror valori (numite scoruri factoriale) se combină liniar pentru a forma scorul total al testului. Această metodă, numită *validare factorială*, se aplică mai ales în cazul testelor care măsoară constructe multidimensionale, când se verifică și dacă variabilele puse în evidență de analiza factorială corespund componentelor constructului.

Se caută semnificația variabilelor măsurate prin scorurile parțiale ale testului, se rețin cele care au legătură cu constructul și se verifică dacă importanța acestora în domeniul constructului este reflectată corect prin ponderile pe care le au măsurile lor în scorul total al testului.

D. Se estimează în ce măsură concluziile desprinse până la acest pas al algoritmului de validare pot fi generalizate, adică pot fi extinse asupra altor populații sau altor ocazii de examinare cu ajutorul testului. Se încearcă să se stabilească dacă interpretările propuse pentru scorurile testului se mențin atunci când se schimbă situația în care se face examinarea, momentul examinării ori evaluatorii.

E. Se cercetează relațiile dintre scorurile testului și diverse alte măsurători psihologice sau observații non-evaluative asupra comportamentului. Această verificare este denumită uneori *validare convergentă și discriminantă*.

Un test are *validitate convergentă* dacă măsoară ceea ce evaluează și alte teste sau variabile care se referă la același construct, adică între scorurile sale și rezultatele respectivelor măsurători există o relație funcțională monotonă (liniară sau nu). Un test are *validitate discriminantă* dacă evaluează altceva decât diverse teste sau variabile despre care se știe că se referă la constructe ce nu au legătură cu constructul măsurat de test. Aceasta înseamnă că scorurile testului nu sunt în relație funcțională monotonă cu rezultatele acelor măsurători.

Verificarea validității convergente și discriminante se realizează conform următorului algoritm:

- (1) Se identifică acele comportamente care sunt legate de constructul măsurat de test.
- (2) Se determină constructele care au legături cu constructul măsurat de test și se pun în evidență acele comportamente legate de ele care pot fi conectate cu constructul măsurat de test.
- (3) Pentru toate comportamentele reținute se apreciază ce fel de corelații ar putea exista între măsurile lor și o măsură bună a constructului care ne interesează (corelații pozitive, negative, mari, mici sau nule).
- (4) Se determină coeficienții de corelație între scorurile la testul studiat și rezultatele unor măsurători (eventual scoruri de teste) efectuate asupra comportamentelor reținute.
- (5) Se compară coeficienții de corelație obținuți cu valorile așteptate.

O metodă utilizată pentru verificarea validității convergente și discriminante a testelor este aceea propusă de D. T. Campbell și D. W. Fiske

(Anastasi, 1976), bazată pe matricea "trăsături multiple - metode multiple" (multitrait-multimethod matrix). Aceasta conține valorile coeficienților de corelație liniară între diverse măsurători ale acelorași constructe, obținute prin intermediul a două sau trei tehnici ori instrumente diferite.

Exemplu

În urma evaluării constructelor A, B și C, despre care se face presupunerea că nu au legături între ele, prin trei metode diferite - "test I", "test II" și "test III" - s-a obținut următoarea matrice de coeficienți de corelație liniară (pentru a se putea urmări mai ușor datele, ea a fost scrisă sub forma unui tabel):

	Test I A B C			Test II A B C			Test III A B C		
Test I A B C	.79								
	.10	.85							
	.03	.34	.75						
Test II A B C	.42			.76					
	.12	.45		.03	.77				
	.06	.19	.56	.12	.17	.83			
Test III A B C	.50			.54			.86		
	.33	.45		.39	.66		.38	.84	
	.25	.40	.58	.29	.37	.63	.35	.41	.85

Observație. Coeficienții cu valori mai mari decât .35 sunt semnificativi la pragul $p=0.05$.

Valorile notate pe diagonala principală sunt coeficienții de fidelitate ai testelor, obținuți prin metoda formelor paralele. De exemplu, pentru testul II, coeficienții de fidelitate sunt 0.76 (pentru constructul A), 0.77 (pentru constructul B) și 0.83 (pentru constructul C).

Valorile de pe diagonalele celorlalte căsuțe sunt coeficienții de corelație între măsurătorile realizate de cele trei teste pentru aceleași constructe. Ei servesc la verificarea validității convergente. De exemplu, pentru constructul A, coeficienții de corelație sunt:

- 0.42 (testul I cu testul II);
- 0.50 (testul I cu testul III);
- 0.54 (testul II cu testul III).

Toți acești coeficienți sunt semnificativi statistic la pragul $p=0.05$, ceea ce denotă o bună validitate convergentă a celor trei teste.

Celelalte valori din matrice reprezintă coeficienții de corelație liniară între constructe diferite, măsurate prin același test sau prin teste diferite. Ei servesc la aprecierea validității discriminante a testelor.

Se observă că testele I și II discriminează între constructele A, B și C (coeficienții de corelație aflați sub diagonala căsuțelor I-I, I-II și II-II sunt nesemnificativi la pragul $p=0.05$), în timp ce testul III nu are validitate discriminantă: coeficienții de corelație liniară între constructele A, B și C, măsurate prin testul III (în căsuța III-III) sau prin testul III și un alt test (în căsuțele I-III și II-III) sunt, în general, semnificativi la pragul $p=0.05$.

În exemplul prezentat, validitatea convergentă și discriminantă a testelor I, II și III a fost ușor de dedus. Dar, în practică, apar situații complexe, greu de interpretat. Uneori corelațiile dintre constructe diferite măsurate prin aceeași metodă sunt mai mari decât corelațiile dintre măsurile aceluiasi construct obținute prin metode diferite (chestionar creion-hârtie, tehnică proiectivă, scală de evaluare, interviu etc.). Aceasta înseamnă că metodele, mai precis unele elemente ale metodelor, au o importanță mai mare în evaluare decât ceea ce se măsoară prin ele.

Mai trebuie reținut că pentru aprecierea validității discriminante a unui test este util să se calculeze, pe lângă coeficientul de corelație dintre test și o altă măsurătoare, și coeficientul de corelație dintre constructele măsurate. Se aplică formula 8.2.2. Întrucât coeficienții de fidelitate iau valori subunitare (valoarea 1 este extrem de rară), corelația constructelor este, în valoare absolută, mai mare decât valoarea absolută a coeficientului de corelație al scorurilor observate. Dacă se obține un coeficient de corelație mare între două constructe care, teoretic, nu au nimic în comun, validitatea discriminantă a testului este necorespunzătoare.

Validarea convergentă a unui test este necesară, dar nu și suficientă pentru a dovedi că testul măsoară ceea ce se dorește. Aceasta, deoarece etichetele atribuite testelor reprezintă trăsăturile sau predispozițiile pe care testele ar trebui să le măsoare. Corelația mare între două teste care poartă aceeași etichetă poate fi cauzată de existența, în cele două teste, a unor itemi asemănători, care nu totdeauna sunt reprezentativi pentru constructul pe care testele intenționează să îl măsoare. Analiza factorială sau analiza de clusteri efectuată asupra itemilor din testele cu aceeași denumire poate ajuta la explicarea corelațiilor dintre scorurile testelor.

F. Se analizează interpretările propuse pentru scorurile testului și sursele de invaliditate a testului, pentru a deduce posibilele consecințe negative ale utilizării testului, atât pe termen scurt, cât și pe perioade mai îndelungate.

Din cele prezentate mai sus se constată că întregul proces de validare relativ la construct a testului constă în formularea de ipoteze științifice (asupra componentelor constructului, asupra relațiilor dintre aceste componente etc.) și în verificarea acestora. Metodele aplicate în validare sunt dependente de ipotezele formulate. Din acest motiv nu există metode valabile pentru validarea relativă la construct a oricărui test.

Trebuie reținut că prin validarea testului relativ la constructul măsurat se validează atât testul cât și ipotezele aflate la baza sa. Validitatea testului și validitatea constructului sunt inseparabile. Când un test nou este elaborat pentru măsurarea unui construct cunoscut, riscul ca validitatea să fie nesatisfăcătoare este mai mare pentru test decât pentru construct. Dar, poate să se constată și necesitatea revizuirii constructului (L.J. Cronbach, după Silva, 1993).

9.3. VALIDITATEA TESTULUI RELATIVĂ LA CONȚINUTUL SĂU .

Acest tip de validitate este analizat la testele care se utilizează pentru a estima "cum acționează o persoană în universul de situații pe care testul intenționează să îl reprezinte" (*Standards for Educational and Psychological Tests*, 1974).

În literatura psihologică există păreri diferite despre ceea ce se urmărește prin validarea testului relativ la conținut. Anastasi (1976), de pildă, consideră că acest tip de validare implică "examinarea sistematică a conținutului testului pentru a determina dacă el acoperă un eșantion reprezentativ din domeniul de comportamente pe care testul în cauză trebuie să îl măsoare". În această accepțiune, itemii testului sunt priviți ca reprezentând un eșantion extras dintr-o populație largă de itemi posibili care definesc ceea ce testul trebuie să măsoare. Dacă eșantionul este reprezentativ pentru populație, atunci se admite că testul are validitate relativă la conținut (Gregory, 1992).

Alți autori înlocuiesc expresia "domeniu de comportamente" cu "mulțimea sarcinilor" sau "mulțimea performanțelor".

R. T. Lennon consideră că validitatea relativă la conținut se referă la răspunsurile subiectului și nu la întrebările testului. Pentru analizarea acestui tip de validitate trebuie luat în considerare nu numai conținutul itemilor testului ci

și procesul utilizat de subiect pentru a ajunge la formularea răspunsului (Silva, 1993). De exemplu, un test având drept domeniu de conținut "cunoștințele de ortografie", format din itemi cu alegere multiplă, poate verifica dacă un individ recunoaște cuvintele scrise corect din punct de vedere ortografic, dar nu și dacă acesta știe să scrie corect după dictare. Deci nu este suficient să se spună că testul evaluează cunoștințele de ortografie.

Adoptând punctul de vedere formulat de Anastasi (1976), analiza validității relative la conținut a unui test impune ca prim pas definirea și descrierea domeniului de conținut al testului. Este esențial ca descrierea să fie cât mai detaliată și să fie precizate clar limitele acestui domeniu.

Pasul următor îl constituie analizarea itemilor testului. Sunt eliminați acei itemi care nu se referă la comportamente cuprinse în domeniul de conținut. Pentru itemii rămași se identifică zonele domeniului de conținut pe care le măsoară. La această operație este bine să participe cât mai mulți experți, care să aprecieze pentru fiecare item dacă se referă la domeniul de conținut al testului și dacă este relevant. O analiză atentă o pretind itemii complecși: aceștia pot să disperseze atenția subiecților și, prin urmare, scorul lor să fie contaminat de alte caracteristici decât aceea care ar trebui să fie măsurată.

Exemplu.

Li se poate cere elevilor să efectueze aceleași operații aritmetice formulând două probleme diferite:

a. Care este cel mai mic număr întreg x pentru care are loc inegalitatea $x - 0.25(x - 1) \geq 5.5$?

b. Un melc urcă pe un stâlp înalt de 5.5 m, pornind de la bază. În fiecare zi el înaintază cu 1 m, iar în timpul nopții alunecă în jos cu 0.25 m. În a câta zi melcul atinge vârful stâlpului?

Cea de a doua problemă le propune subiecților o sarcină mai complexă decât prima, întrucât le pretinde să citească un text mai lung și să dea dovadă de atenție, să observe că întrebarea se referă la ziua în care este "atins" vârful stâlpului. Răspunsul corect este "ziua a șaptea", chiar dacă după ce ajunge în vârful stâlpului, în ziua a șaptea, melcul coboară cu 0.25m în cursul nopții. Dacă problema a doua este inclusă într-un test de raționament aritmetic pentru elevi, este posibil ca scorul său să fie alterat de variabila irelevantă "deprinderea de a citi".

Din acest motiv este important să se verifice pentru fiecare item dacă nu cumva necesită un comportament care nu aparține domeniului de conținut al testului.

Ultima etapă a validării constă în compararea structurii testului cu structura domeniului de conținut. Se verifică dacă sunt acoperite toate zonele

domeniului de conținut și dacă itemii sunt repartizați proporțional cu mărimea și importanța zonelor.

Aprecierea generală privind validitatea testului relativă la conținutul său este subiectivă. Ea nu apelează la metode statistice, ci se bazează doar pe raționamente. Din acest motiv, validarea relativă la conținut este denumită în literatură și *validare logică* sau *validare rațională* (Thorndike, Hagen, 1961; Anastasi, 1976).

Totuși, în cazul testelor care verifică deprinderi sau cunoștințe însușite în cadrul unui program de instruire există recomandarea de a lua în considerare, la aprecierea validității relative la conținut a testului, și rezultatul comparării statistice a scorurilor obținute la test de aceleași persoane, înainte și după acest program (Anastasi, 1954). Într-adevăr, dacă există o suprapunere între conținutul itemilor testului și conținutul programului de instruire și dacă itemii sunt bine construiți, atunci este de așteptat ca scorurile finale să fie mult mai mari decât cele inițiale, deci media diferențelor de scor să fie pozitivă și să difere semnificativ de zero.

Validarea relativă la conținut se impune în cazul testelor utilizate la măsurarea unui atribut ce nu poate fi exprimat printr-un construct (Murphy, Davidshofer, 1991). De exemplu, putem descrie destul de detaliat domeniul de conținut pentru "cunoștințe de psihometrie", dar este foarte dificil să găsim comportamentele prin care această caracteristică să poată fi observată.

Validitatea relativă la conținut se studiază mai ales în cazul testelor de cunoștințe, atunci când nu există nici un criteriu extern potrivit pentru a analiza validitatea relativă la criteriu.

Trebuie menționat că validarea relativă la conținut nu permite să se tragă nici o concluzie despre calitatea măsurătorii realizate de test (Murphy, Davidshofer, 1991). De asemenea, ea nu furnizează suficiente informații pentru a putea interpreta corect scorurile testului. Nu se poate afirma că, dacă testul are validitate relativă la conținutul său, atunci în privința atributului măsurat de test, un subiect care a realizat un scor mai mare îi este superior unuia care a obținut un scor mai mic.

Deci, operațiile cuprinse sub denumirea de "validarea testului relativă la conținutul său" nu reprezintă o validare, în sensul definiției date de *Standards for Educational and Psychological Tests* din 1985. Ele realizează doar analiza testului în privința clarității, a reprezentativității și a relevanței conținutului său.

9.4. VALIDITATEA TESTULUI RELATIVĂ LA UN CRITERIU

9.4.1. Metode de cercetare a validității relative la criteriu

Validarea relativă la criteriu se referă la gradul în care deducțiile făcute pornind de la scorurile testului concordă cu cele bazate pe valorile unei alte măsurări, numită *criteriu*. Validarea relativă la criteriu conține proceduri care compară scorurile testului cu evaluări, clasificări sau măsurători (mentale ori comportamentale) ale aceluiași subiecți.

Pentru fiecare dintre criteriile utilizate la validare, este necesar ca în manualul testului să se explice motivele alegerii sale și să i se facă o descriere amănunțită (*Standards for Educational and Psychological Tests*, 1985).

Aprecierea validității relative la criteriu se bazează, de obicei, pe valorile unui *coeficient de validitate a testului*, care exprimă numeric rezultatul comparării scorurilor testului cu valorile criteriului.

Coeficientul de validitate al unui test, conform definiției din majoritatea lucrărilor de teoria testelor, este egal cu coeficientul de corelație liniară între scorurile testului și rezultatele unei măsurători considerate drept variabilă criteriu. Unii autori (Lord, Novick, 1968) înțeleg prin coeficientul de validitate valoarea absolută a coeficientului de corelație liniară; alți autori (Cureton, 1955) impun calculul coeficientului de corelație între scorurile observate la test și scorurile reale ale variabilei criteriu.

În literatura de specialitate există și recomandarea de a calcula coeficientul de validitate folosind alți indici în locul coeficientului de corelație liniară. De exemplu, Cronbach (1966) permite calcularea validității relative la criteriu cu ajutorul coeficientului de corelație a rangurilor.

De fapt, în majoritatea cazurilor, testele și criteriile realizează doar măsurări pe scală ordinală, nu pe scală de interval. Din acest motiv, calculul coeficientului de corelație liniară între scorurile testului și valorile criteriului nu este corectă. Ținând cont de faptul că prin compararea scorurilor testului cu valorile criteriului se realizează, de fapt, compararea ierarhiilor produse în rândul subiecților de test și de criteriu, pentru studiul validității relative la criteriu se pot utiliza toți coeficienții care servesc la compararea a două clasamente prezentați în Anexa I.

Se apreciază că un test are validitate bună atunci când coeficientul de validitate este mare.

Uneori se poate îmbunătăți valoarea coeficientului de validitate prin gruparea scorurilor testului, urmată de transformarea valorilor acestora (Albu, Pitariu, 1994 și 1995a).

O altă posibilitate de cercetare a validității unui test relativă la criteriu utilizează două grupe contrastante de subiecți, formate pe baza valorilor criteriului: una este compusă din indivizi cu rezultate foarte bune, cealaltă din persoane cu rezultate foarte slabe. Se compară statistic mediile scorurilor obținute la test de persoanele din cele două grupe și se apreciază că testul este valid relativ la criteriu dacă mediile diferă semnificativ, la un prag de probabilitate fixat (Anastasi, 1976). Această metodă se aplică mai ales atunci când variabila criteriu este calitativă - deci nu se poate calcula coeficientul de corelație liniară - și atunci când criteriul este multidimensional, având mai multe componente cantitative și/sau calitative.

Unii autori (Hammond, 1995) recomandă ca validarea unui test relativă la criteriu să utilizeze mai multe variabile criteriu, dintre care unele nu au legătură cu ceea ce testul măsoară sau prezice. Acestea servesc la verificarea *specificității* testului. Este de dorit ca testul să nu coreleze cu ele.

În cazul testelor utilizate pentru măsurare, validarea relativă la criteriu este o componentă a validării relative la constructul măsurat. Ea apreciază "gradul în care rezultatele furnizate de test corelează cu rezultatele altui instrument de evaluare despre care se presupune că măsoară același construct sau unul similar" (S. N. Haynes, citat după Silva, 1993). Evident, calculul coeficientului de corelație trebuie să fie precedat de analiza constructelor măsurate de test și de criteriu.

Chiar și validarea factorială utilizată pentru cercetarea validității relative la construct a unui test este, de fapt, o validare relativă la criteriu (Anastasi, 1954). Criteriul este reprezentat, în acest caz, de factorul care corespunde constructului măsurat de test. Una dintre condițiile impuse pentru a putea aprecia că testul este valid este ca itemii săi să aibă saturații mari în factorul respectiv. Or, saturația unui item într-un factor este valoarea coeficientului de corelație liniară între scorurile itemului și scorurile factorului.

În cazul testelor utilizate pentru a se lua decizii, validarea relativă la criteriu urmărește dacă deciziile luate folosind scorurile testului coincid cu cele care au la bază valorile unei variabile criteriu, despre care se presupune că permite să se ia decizii corecte. De exemplu, decizia poate consta în acceptarea sau respingerea candidaților care concurează pentru un post. Se dovedește că testul este valid relativ la acest criteriu dacă se constată că indivizii acceptați (pe baza unor considerente care nu includ rezultatele testului) sunt tocmai cei care au scoruri mari la test, iar cei respinși sunt cei cu scoruri mici.

Validarea testelor utilizate pentru predicție are în vedere gradul în care scorurile testului - numit în acest caz *predictor* - pot fi utilizate pentru a se deduce performanța la o variabilă numită *criteriu*, independentă de test (Guion, 1974). Trebuie menționat că o aceeași variabilă poate să joace rol de predictor într-un

caz (de exemplu, performanța școlară ca predictor pentru performanța profesională) sau rol de criteriu în alt caz (de exemplu, aceeași performanță școlară ca și criteriu pentru un test de inteligență).

Există două strategii utilizate în mod obișnuit pentru aprecierea validității relative la criteriu a testelor:

- validarea concurentă;
- validarea predictivă.

Validarea concurentă se aplică îndeosebi pentru testele utilizate la obținerea unor diplome sau licențe, care măsoară cunoștințele sau deprinderile persoanelor, și pentru testele care servesc la stabilirea unui diagnostic clinic. Se cercetează validitatea concurentă atunci când se dorește să se cunoască în ce măsură scorurile testului estimează pozițiile ocupate în prezent de subiecți la un criteriu relevant. De exemplu, pentru un test de depresie, validitatea concurentă oferă informații despre gradul de coincidență dintre concluziile desprinse din interpretarea scorurilor testului și părerile psihologilor clinicieni sau cele ale psihiatrilor, relative la aceiași indivizi.

Pentru cercetarea validității concurente este necesară obținerea scorurilor la criteriu aproximativ în același timp cu cele ale testului.

Validarea predictivă se aplică atunci când scorurile testului sunt folosite pentru a prezice valorile pe care o anumită variabilă le va avea într-un viitor mai mult sau mai puțin apropiat. Asemenea teste se utilizează de obicei la admiterea în școli sau la selecția profesională, pentru a-i alege pe acei candidați care au șanse mari ca la sfârșitul cursurilor sau după ocuparea postului să obțină performanțe bune.

De exemplu, pentru estimarea validității predictive a unui test utilizat în selecția profesională se calculează coeficientul de corelație liniară între scorurile testului, administrat cu ocazia finalizării unui curs de pregătire profesională, și măsurile performanțelor obținute de aceleași persoane după un anumit timp, în care au avut posibilitatea să aplice cunoștințele respective.

Trebuie subliniat faptul că validitatea predictivă se cercetează folosind loturi neselectate. Aceasta înseamnă că toate persoanele cărora li s-a administrat testul vor urma același tratament și vor fi prezente la evaluarea variabilei criteriu. În acest fel se asigură ca în lotul de subiecți care servește la compararea testului cu criteriul să existe atât scoruri mari la test, cât și scoruri mici. Nu totdeauna se poate realiza așa ceva. Este greu de acceptat, de exemplu, ca să fie angajate toate persoanele care se prezintă la selecția profesională sau ca decizia referitoare la angajare să se ia la întâmplare.

Atunci când testul va fi utilizat pentru a face predicții, este necesar ca, alături de coeficientul de validitate al testului să se determine și ecuația de regresie

valoarea criteriului = $f(\text{scor la test})$

Aceasta permite să se estimeze rezultatul pe care îl va obține la criteriu fiecare persoană, cunoscând scorul său la test. De exemplu, dacă relația dintre media de absolvire a facultății, F , și media obținută la admiterea în facultate, A , poate fi scrisă sub forma

$$F = 0.3 + 0.9A$$

atunci este de așteptat ca o persoană care a fost admisă în facultate cu media 8.5 să termine facultatea cu media

$$F = 0.3 + 0.9 \times 8.5 = 7.95$$

9.4.2. Problema restrângerii mulțimii

Deosebirea esențială dintre cele două tipuri de validări relative la criteriu nu constă în lungimea intervalului de timp dintre administrarea testului și înregistrarea rezultatelor criteriului, ci în faptul că validarea concurentă utilizează un eșantion de populație selecționat (personal angajat deja sau studenți admiși în facultăți). Apare aici următoarea problemă, denumită *restrângerea mulțimii*, care face ca validarea concurentă să nu poată furniza totdeauna informații corecte asupra capacității predictive a testului:

Lotul de persoane utilizat la calculul coeficientului de validitate provine din populația asupra căreia ar fi trebuit să se facă predicția, dar până în momentul când se înregistrează valorile variabilei criteriu el a fost supus deja la două selecții.

(1) Pe de o parte, au fost reținuți din eșantionul inițial doar cei care au obținut rezultate bune la criteriu (de exemplu, cei care au avut note mari la admiterea în facultate). Dacă acest criteriu măsoară același lucru ca și testul, atunci persoanele rămase sunt cele care vor obține la test scoruri mai mari decât ar fi realizat cei respinși.

(2) Pe de altă parte, au fost eliminați pe parcurs cei cu performanțe școlare sau profesionale slabe, adică cei cu valori mici la variabila criteriu utilizată în studiul validității. Este posibil să fi fost înlăturați prin promovare, și cei cu valori foarte mari la criteriu.

Deci, pentru lotul considerat, atât mulțimea scorurilor testului cât și mulțimea valorilor criteriului reprezintă restrângeri ale mulțimilor corespunzătoare care s-ar fi obținut pentru întreaga populație. În consecință, coeficientul de corelație liniară calculat pe baza persoanelor selecționate diferă de cel care ar fi rezultat lucrând cu întreaga populație și care ar fi fost util pentru aprecierea valorii predictive a testului.

Exemplu.

Să considerăm un lot format din 105 persoane care au răspuns la un test de capacitate de organizare, ale cărui scoruri posibile sunt 1, 2, 3, 4 și 5. Fiecare subiect a fost evaluat în privința capacităților sale manageriale, primind o notă cuprinsă între 1 și 5.

Rezultatele obținute de persoanele din lot sunt notate în următorul tabel de frecvențe:

Scorul testului	Valoarea criteriului				
	1	2	3	4	5
1	10	2	2	1	2
2	2	8	4	0	2
3	0	6	18	3	0
4	2	1	4	11	6
5	1	2	4	7	7

Pe întregul lot, coeficientul de corelație liniară între test și criteriu are valoarea 0.520, semnificativă la pragul $p=0.001$.

Dacă din lot au fost reținuți doar cei 45 de subiecți care au realizat la test cel puțin 4 puncte, atunci coeficientul de corelație liniară între test și criteriu este egal cu 0.026, fiind nesemnificativ la pragul $p=0.05$.

Dacă din lot au fost eliminate persoanele care au fost evaluate cu note mai mici decât 4, atunci pentru cele 39 de persoane rămase coeficientul de corelație liniară dintre test și criteriu ia valoarea -0.095, fiind nesemnificativ la pragul $p=0.05$.

Pentru 31 de persoane care au primit cel puțin nota 4 la evaluare și au obținut cel puțin 4 puncte la test coeficientul de corelație între test și criteriu este egal cu 0.148, fiind nesemnificativ la pragul $p=0.05$.

9.4.3. Interpretarea coeficienților de validitate

Hull (Guilford, 1965) a impus condiția ca un test să fie utilizat în practică doar dacă are un coeficient de validitate mai mare decât 0.45. Această cerință este însă greu de realizat. Un test bun rareori poate da, cu un criteriu

important, un coeficient de corelație mai mare decât 0.5 (Murphy, Davidshofer, 1991).

În experiența obișnuită, valoarea coeficientului de validitate al testelor utilizate pentru a face predicții este cuprinsă între 0 și 0.60; în majoritatea cazurilor fiind situată în partea inferioară a intervalului [0, 0.60] (Guilford, 1965).

Constatarea că, în general, comportamentul aceleiași persoane variază de la o situație la alta a condus la concluzia că valoarea coeficientului de validitate al testelor de personalitate administrate în scopul predicției unei acțiuni nu poate fi mai mare decât 0.4 (Funder, 1983).

O posibilitate de a stabili dacă validitatea relativă la criteriu a unui test este sau nu acceptabilă pornește de la eroarea maximă acceptată în estimarea valorilor criteriului și se bazează pe formula de calcul a erorii standard de estimare (Gregory, 1992):

$$ES_{est} = \sigma_c \sqrt{1 - r_{ct}^2} \quad (9.4.3.1)$$

unde:

ES_{est} = eroarea standard de estimare, care reprezintă marginea superioară a erorii de estimare a valorilor criteriului prin ecuația de regresie liniară;
 σ_c = abaterea standard a valorilor criteriului;
 r_{ct} = coeficientul de corelație liniară între test și criteriu;

ES_{est} ajută să se răspundă la întrebarea "Cât de precis poate fi prezisă performanța la criteriu, pornind de la scorurile testului?"

Presupunând că repartiția de frecvențe a valorilor criteriului urmează legea normală și că ecuația de regresie a criteriului (c) față de test (t) este:

$$c = a + bt$$

există șanse de 95% ca valoarea criteriului pentru o persoană care a obținut scorul t_0 la test să se afle în intervalul:

$$[a + bt_0 - 1.96ES_{est}, a + bt_0 + 1.96ES_{est}]$$

Cu cât coeficientul de validitate al testului, r_{ct} , este mai mare, cu atât eroarea standard de estimare este mai mică, iar predicția realizată cu ajutorul testului este mai precisă.

9.4.4. Factori care afectează valoarea coeficienților de validitate

Valoarea coeficientului de validitate al unui test este influențată nu numai de test, ci și de criteriul utilizat și de populația căreia i se administrează

testul. Amintim câțiva dintre factorii care afectează validitatea testului (adaptare după Aiken, 1994 și Warr, 1997).

A. Unele caracteristici ale populației, precum sexul, vârsta sau chiar anumite trăsături de personalitate, acționează ca variabile moderatoare, modificând relația dintre criteriu și test. Subiecții se împart în grupe, în funcție de valorile variabilei moderatoare, iar valoarea coeficientului de validitate al testului se poate schimba de la o grupă la alta.

În general, coeficienții de validitate sunt mai mici în grupele mai omogene, adică în grupele în care scorurile testului sau valorile criteriului variază puțin.

Atunci când valoarea coeficientului de validitate al unui test este scăzută, iar ceea ce se cunoaște despre test și despre criteriu nu oferă o explicație, trebuie să se verifice dacă nu există o variabilă moderatoare care se interpune între criteriu și test.

Întrucât validitatea unui test este dependentă și de grupul de subiecți, ori de câte ori se constată că un test nou este valid față de un anumit criteriu, este necesar să se efectueze o validare suplimentară (cross-validation), folosind același criteriu, dar un alt eșantion de subiecți, proveniți din aceeași populație ca și primul. Doar dacă și în acest caz testul se dovedește a fi valid, se poate recomanda utilizarea sa în practică.

B. La fel ca și fidelitatea, validitatea relativă la criteriu este influențată de lungimea testului.

Dacă testul este format din itemi paraleli și se modifică lungimea testului prin eliminarea unor itemi sau prin adăugarea de itemi paraleli, atunci coeficientul de validitate al testului obținut, relativ la același criteriu pentru care a fost calculat coeficientul de validitate al testului original, este dat de formula (Burisch, 1997):

$$r_k = \frac{r_0 \sqrt{k}}{\sqrt{1 + (k-1)r_u}} \quad (9.4.4.1)$$

unde:

r_k = coeficientul de validitate al testului nou;
 r_0 = coeficientul de validitate al testului original;
 k = raportul dintre numărul de itemi în testul nou și numărul de itemi în testul original;
 r_u = coeficientul de fidelitate al testului original.

Se constată că prin adăugarea de itemi paraleli la un test validitatea testului crește.

C. Validitatea testului este dependentă și de validitatea criteriului ca și măsură a variabilei la care se referă testul.

Dacă valorile variabilei criteriu sunt influențate de scorurile testului, se spune că a avut loc o *contaminare a criteriului*. Aceasta are ca efect creșterea artificială a validității relative la criteriu a testului.

În majoritatea cazurilor, contaminarea criteriului se produce atunci când testul și criteriul au itemi comuni sau în acele situații când valorile criteriului sunt rezultatul unor evaluări făcute de experți, iar aceștia cunosc scorurile obținute de către subiecți la test și, intenționat sau nu, sunt influențați de ele în formularea aprecierilor.

D. Un alt factor care influențează mărimea coeficientului de validitate este frecvența relativă în populație a persoanelor care prezintă caracteristica la care se referă testul (base rate). Atunci când comportamentul pe care testul îl măsoară sau îl prezice este foarte rar întâlnit (precum psihoticismul sau suicidul), validitatea relativă la criteriu este de obicei scăzută.

E. Validitatea unui test este dependentă de fidelitatea sa. Într-adevăr, dacă $\alpha \in M$ și $b \in M$ sunt două măsurători paralele, iar $c \in M$ este o măsurătoare arbitrară, atunci are loc următoarea inegalitate:

$$\rho(X_a, X_c) \leq \sqrt{\rho(X_a, X_b)} \quad (9.4.4.2)$$

Aceasta se interpretează astfel: coeficientul de validitate al unui test $\alpha \in M$ relativ la orice criteriu $c \in M$ nu poate depăși indicele de fidelitate al testului α . Înseamnă că, dacă un test nu este fidel, validitatea sa relativ la orice criteriu nu poate fi mare. Dar, o valoare mare a indicelui de fidelitate nu garantează o validitate mare.

F. Uneori, coeficientul de validitate ia o valoare scăzută nu din cauza testului, ci a criteriului utilizat. Așa se întâmplă atunci când:

- criteriul are o fidelitate scăzută;
- intervalul de valori pentru criteriu este restrâns, neacoperind decât o parte din mulțimea valorilor posibile;
- repartiția de frecvențe a valorilor criteriului nu este normală;

- valorile criteriului nu sunt acordate sau notate corect, fie din cauza unor factori care țin de evaluator (oboseală, neatenție, incompetență sau părtinire), fie pentru că respectivul criteriu este dificil de măsurat.

9.5. CÂTEVA IDEI IMPORTANTE DESPRE VALIDITATEA TESTELOR

La alegerea strategiilor de validare a unui test, precum și la interpretarea coeficienților de validitate, este util să se ia în considerare și următoarele păreri (adaptare după Messick, 1981 și Hammond, 1995):

A. Validitatea unui test este dependentă de scopul pentru care acesta este folosit și de populația căreia i se administrează testul. Nici un test nu este valid în orice situație de utilizare. Concluziile desprinse dintr-un studiu privind validitatea unui test sunt influențate și de metodele aplicate în cercetare.

B. Există diverse metode pentru validarea testelor: analiza de itemi, analiza factorială, calculul coeficienților de corelație dintre test și diverse criterii etc. Alegerea strategiei de validare pentru un anumit test trebuie să fie determinată de scopul pentru care se va utiliza testul și de cerințele formulate de cel care a solicitat examinarea psihologică în cadrul căreia se intenționează să se administreze testul respectiv.

C. Validitatea unui test nu poate fi estimată printr-un singur coeficient, ci se deduce din acumularea dovezilor empirice și conceptuale.

Standards for Educational and Psychological Tests din 1974 atrag atenția că "validitatea nu se măsoară, ci se deduce". Manualul testului trebuie să conțină valorile mai multor coeficienți de validitate, calculați pentru utilizări diferite ale testului sau pentru populații diferite. În cazul unei aplicări particulare a testului nu se va lua în considerare un singur coeficient, ci prin analiza lor globală se va aprecia dacă utilizarea testului este "adecvată" sau "nesatisfăcătoare".

Pentru a sublinia ideea că este greșit să se spună "testul are validitate (de un anumit tip)", *Standards for Educational and Psychological Tests* din 1985 modifică terminologia utilizată în legătură cu validitatea. Astfel, ele înlocuiesc denumirile cunoscute ale "tipurilor" de validitate cu următoarele "etichete" atribuite strategiilor de validare: "*dovezi ale validității relative la conținut*", "*dovezi ale validității relative la construct*" și "*dovezi ale validității*

relative la criteriu". Prin urmare, faptul că scorurile testului folosit ca predictor corelează liniar semnificativ, la un prag de probabilitate fixat, cu valorile unui criteriu nu reprezintă o evaluare a validității relative la criteriu și nu permite să se afirme că "testul are validitate relativă la criteriu", ci doar constituie una dintre dovezile validității relative la criteriu a testului.

D. Strategiile de validare se aplică încă din faza de construcție a testului. Ele se repetă până când se obține un test acceptabil în ceea ce privește lungimea, fidelitatea, validitatea și ușurința în administrare, cotare și interpretare. Procedeele de validare sunt următoarele:

a. În cazul testelor care măsoară constructe, se aleg și se introduc în test itemi din două categorii: unii considerați a fi reprezentativi pentru domeniul constructului și unii despre care se presupune că nu au nici o legătură cu constructul. După administrarea testului sunt reținuți doar acei itemi care se aseamănă între ei în privința răspunsurilor obținute de la subiecți. Se analizează conținutul acestora pentru a se vedea dacă, într-adevăr, ei exemplifică teoria constructului ce trebuie măsurat și dacă se deosebesc de itemii considerați a fi nerelevanți pentru construct.

b. În cazul testelor care vor fi validate relativ la conținutul lor, se introduc în test doar itemi care acoperă domeniul de conținut al testului. După administrarea testului, se rețin itemii la care s-au obținut răspunsuri asemănătoare. În măsura posibilității, se verifică dacă ei corelează cu un criteriu care se referă la același domeniu de conținut sau dacă răspunsurile obținute la acești itemi se deosebesc semnificativ între două grupe contrastante de subiecți.

c. În cazul testelor care vor fi utilizate pentru a face predicții sau a lua decizii, pentru care se poate alege un criteriu de validare, după administrarea testului se rețin doar itemii care se comportă față de acest criteriu așa cum se dorește să se comporte întregul test.

E. Validarea unui test este un proces care trebuie repetat mereu pe parcursul perioadei în care acesta este utilizat, întrucât datele care se acumulează, provenite din observații clinice și din diverse proiecte de cercetare, permit ca interpretarea propusă inițial pentru scorurile testului să fie îmbunătățită și rafinată.

Ori de câte ori se face o modificare importantă într-un test în ceea ce privește formatul, limbajul, conținutul sau instrucțiunile de administrare, testul va fi validat din nou.

CAPITOLUL 10 DIFICULTATEA ȘI SENSIBILITATEA UNUI TEST PSIHOLOGIC

Aceste proprietăți sunt discutate, de obicei, atunci când se verifică dacă un test este adecvat unui lot de persoane. Elementele analizate în acest scop sunt, în majoritatea cazurilor, forma repartiției de frecvențe și valorile indicilor statistici (media, dispersia) pentru scorurile realizate la test de lotul respectiv sau de un eșantion asemănător acestuia.

10.1. DIFICULTATEA UNUI TEST PSIHOLOGIC

Se consideră că un test este **dificil** dacă un număr mare de persoane obțin scoruri mici, este **ușor** dacă majoritatea scorurilor realizate sunt mari și este **potrivit** lotului dacă scorurile sunt repartizate pe un interval de valori destul de mare, de obicei fiind grupate în jurul mediei. În primul caz, repartiția de frecvențe pentru scorurile obținute la test este asimetrică la stânga (are forma literei *i*) (figura 10.1.1a), în cel de-al doilea caz este asimetrică la dreapta (are forma literei *j*) (figura 10.1.1b), iar în ultimul caz este simetrică față de verticala ridicată în punctul corespunzător mediei scorurilor (figura 10.1.1c). De obicei se apreciază că un test este *adecvat unei populații* dacă repartiția scorurilor sale urmează legea normală.

O asemenea analiză a dificultății este posibilă nu numai pentru testele de eficiență (de aptitudini și de cunoștințe), ci și pentru chestionarele de personalitate, deși în cazul acestora din urmă termenul "dificultate" nu pare potrivit.

Dificultatea unui test, apreciată pe baza repartiției de frecvențe a scorurilor rezultate în urma unei administrări, este determinată atât de conținutul testului, cât și de caracteristicile lotului de subiecți.

Atunci când itemii care alcătuiesc testul sunt puternic corelați între ei, fiecare persoană va da răspunsuri cotate identic sau asemănător, pentru că îi va rezolva pe toți corect (respectiv, greșit) sau va alege răspunsul care corespunde valorii mari (respectiv, mici) a variabilei măsurate de test. Prin urmare, în orice lot de indivizi se vor obține doar scoruri foarte mari și/sau foarte mici.

Pentru un chestionar de personalitate cu itemi dihotomici este posibil ca populația căreia i se administrează testul să se împartă în două grupe: unii care la majoritatea itemilor dau răspunsul corespunzător scorului 1 și alții care

aleg răspunsul cotat cu 0. Se obține, deci, o repartiție bimodală a scorurilor (figura 10.1.1.d). O asemenea situație poate să apară, de pildă, la o scală de feminitate/ masculinitate, când răspunsurile la itemi se diferențiază în funcție de sex.

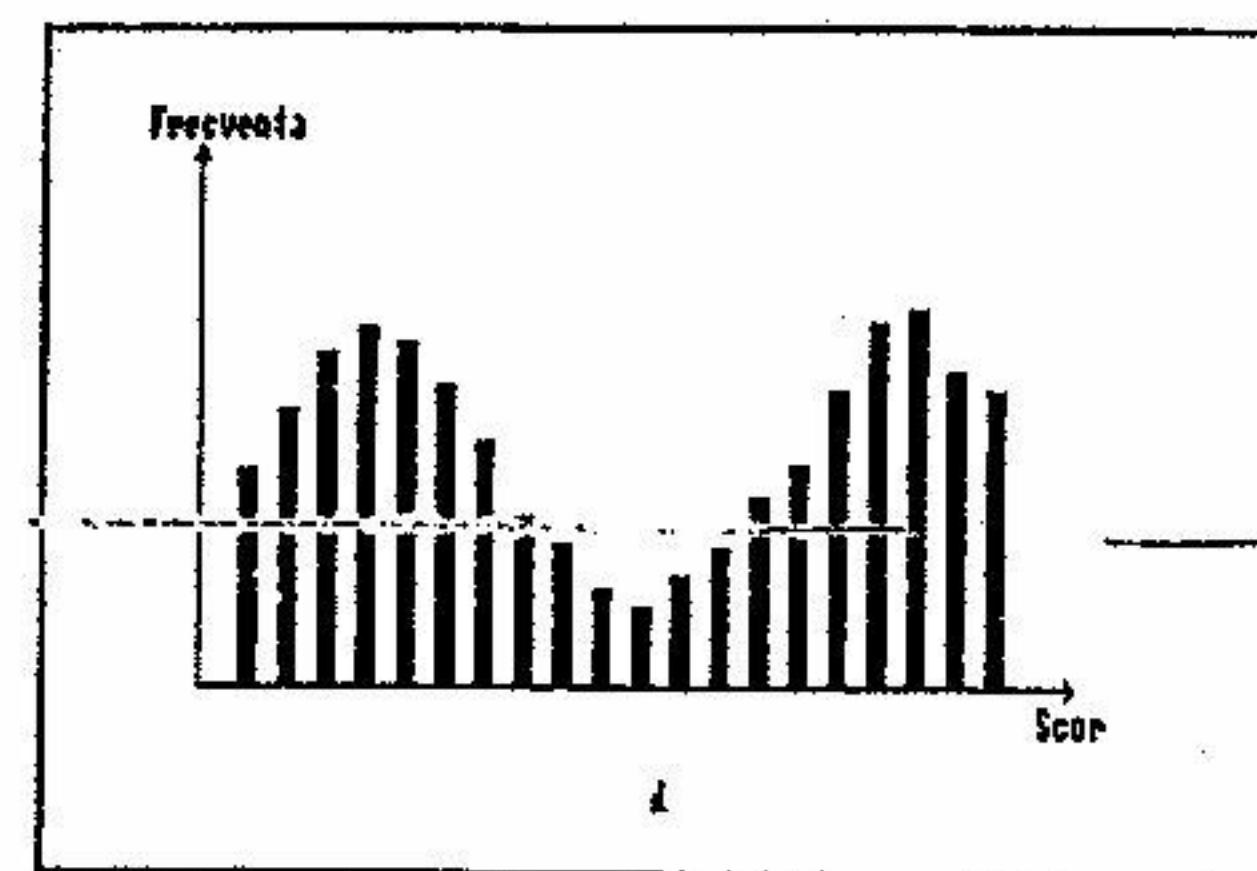
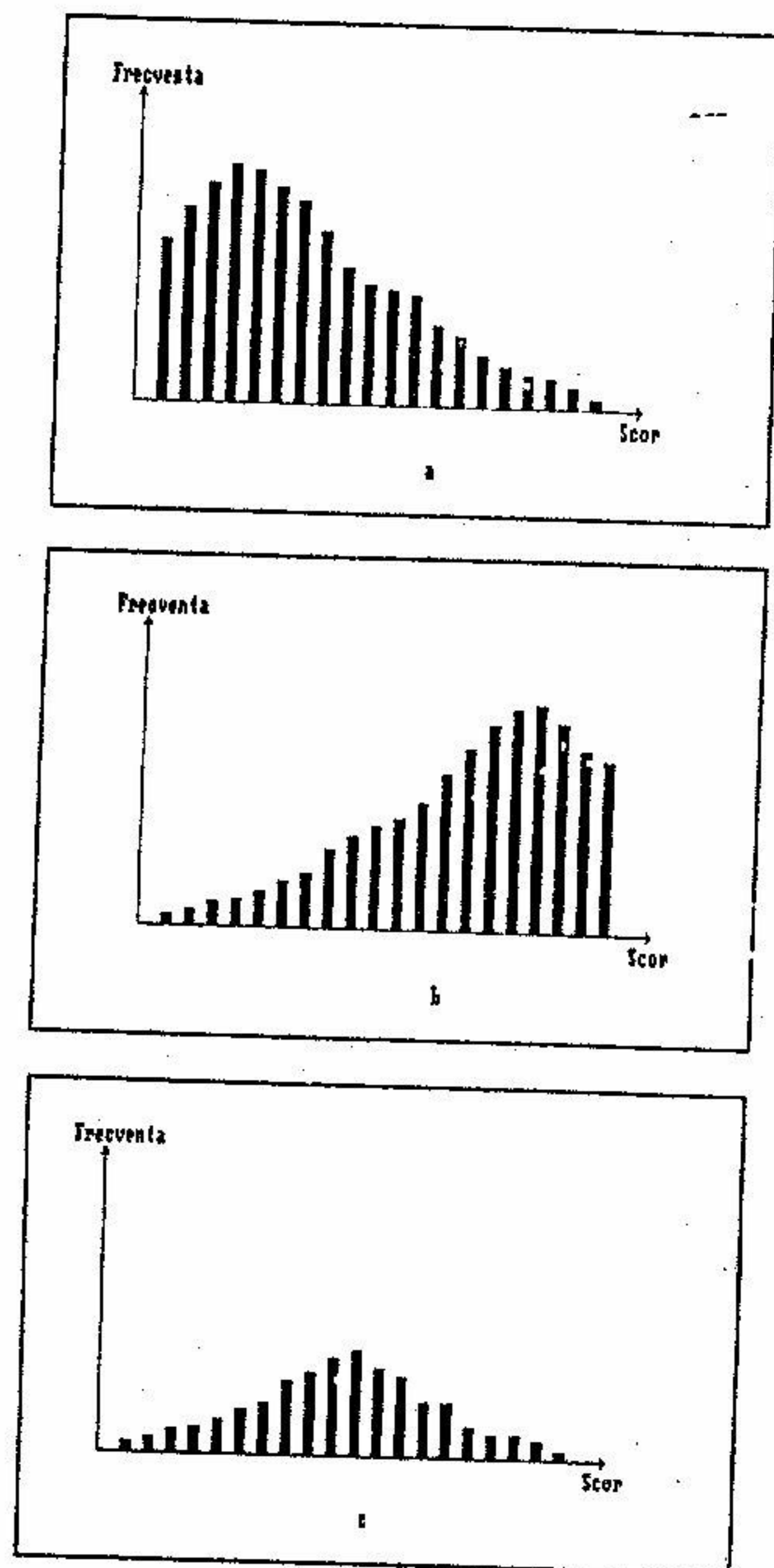


Figura 10.1.1.

Diverse repartiții de frecvențe ale scorurilor unor teste

Testele centrate pe criteriu produc, de obicei, repartiții asimetrice sau bimodale. Ele conțin itemi care verifică existența unor priceperi sau cunoștințe absolut necesare unui anumit scop (promovarea unui program de instruire, ocuparea unui post etc.). De exemplu, ele pot solicita rezolvarea unor exerciții de aritmetică simplă, la sfârșitul clasei întâi, sau executarea unei piese dificile, pentru ocuparea unui post care pretinde o calificare superioară. Utilizarea acestor teste nu urmărește ierarhizarea persoanelor, ci împărțirea lor în două categorii: "capabile" și "incapabile" pentru scopul fixat.

Un test de cunoștințe, administrat aceluiași indivizi în momente diferite, poate produce repartiții total deosebite ale scorurilor.

Exemplu (adaptare după G. de Landsheere, 1975).

Să considerăm un test de cunoștințe de statistică, destinat studenților anului I de la Psihologie.

La începutul anului, majoritatea studenților nu cunosc principalele formule statistice utilizate la prelucrarea datelor experimentale; unii știu să calculeze media, mediana și dispersia și doar câțiva au lucrat cu coeficienți de corelație. Rezultatele care s-ar obține la test ar avea curba de repartiție asimetrică la stânga.

După trecerea examenului de "Psihologie experimentală și analiza datelor", majoritatea studenților cunosc și aplică bine formulele statistice; unii au dificultăți la aplicarea câtorva formule și numai foarte puțini nu cunosc formulele. Același test de cunoștințe ar avea curba de repartiție a scorurilor asimetrică la dreapta.

În intervalul de timp delimitat de începutul anului și sfârșitul sesiunii există, probabil, un moment în care cei mai conștiincioși au învățat mai mult, cei mai puțin interesați de statistică au rămas în urmă și, dacă s-ar administra testul de cunoștințe, curba de repartiție a scorurilor ar fi asemănătoare curbei normale. Schematic, evoluția este prezentată în figura 10.1.2.

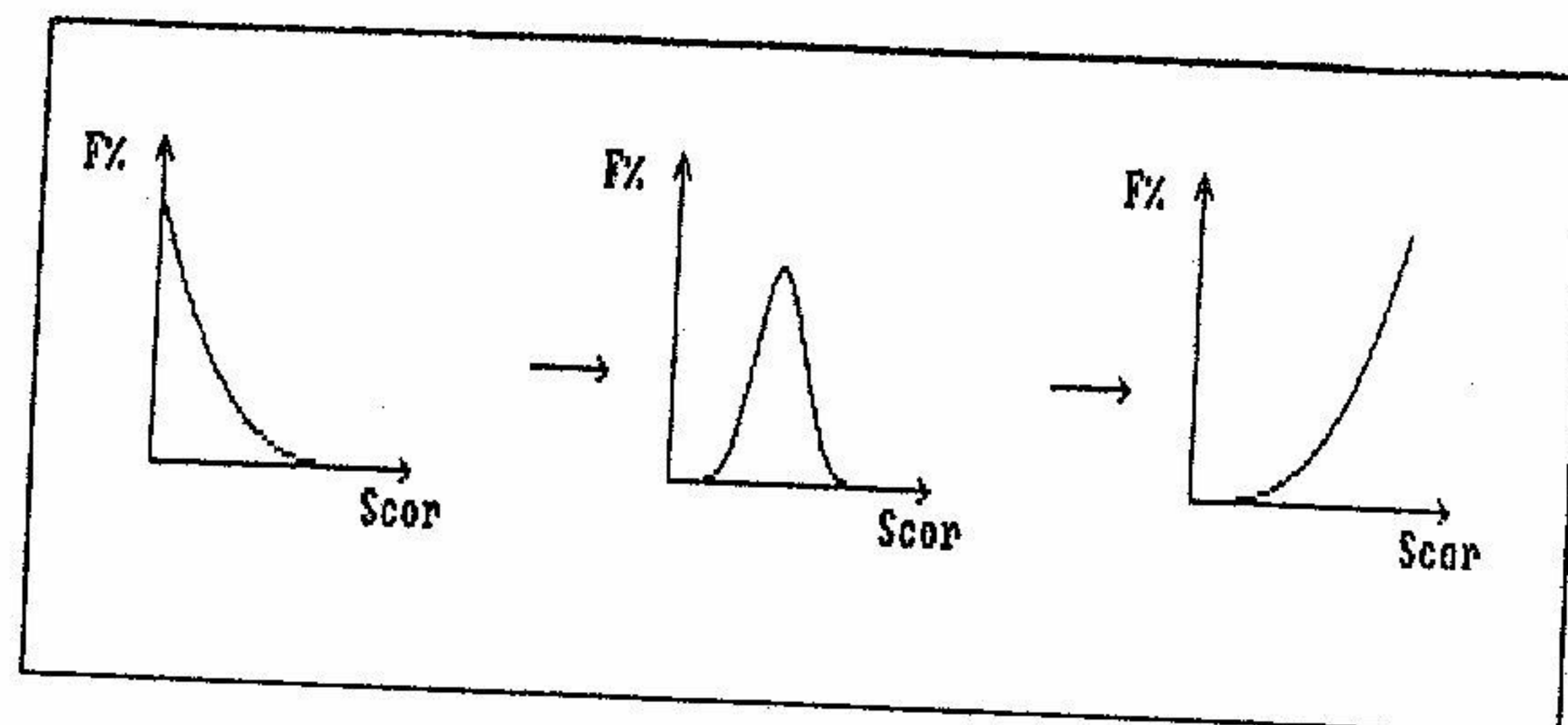


Figura 10.1.2
Curbele de repartiție pentru scorurile aceluiași test de cunoștințe, în trei momente diferite

10.2. SENSIBILITATEA UNUI TEST PSIHOLGIC

Puterea de discriminare sau sensibilitatea unui test este dată de capacitatea acestuia de a face distincție între persoanele examinate, în privința trăsăturii de personalitate sau a aptitudinii măsurate (Lafon, 1973), respectiv de a produce scoruri diferite pentru subiecții care se deosebesc între ei în privința caracteristicii măsurate de test (Kline, 1993).

Piéron (1952) consideră că puterea de discriminare a unui test este dată de numărul claselor pe care le permite în cadrul unui grup. Cu cât numărul valorilor de scor obținute de persoanele unui lot este mai mare, cu atât testul diferențiază mai bine între indivizi. Dar, numărul de valori de scor rezultate în urma administrării testului depinde atât de test (de itemii componenți), cât și de caracteristicile lotului (cât de mari sunt deosebirile dintre persoane în privința variabilei măsurate sau prezise de test).

Atunci când cota brută a testului se obține prin însumarea cotelor itemilor, este de așteptat ca un test alcătuit din mulți itemi să furnizeze mai multe valori de scor decât unul format dintr-un număr redus de itemi. Dar, așa cum s-a arătat mai sus, dacă itemii sunt puternic intercorelați, există riscul de a se obține doar scoruri foarte mari și/sau foarte mici.

Ausubel și Robinson (1981) atrag atenția asupra dependenței dintre sensibilitatea testului și forma repartiției scorurilor.

O repartiție a rezultatelor apropiată ca formă de curba normală, de exemplu, asigură o discriminare maximă la ambele capete ale scalei (unde sunt risipite puține scoruri) și o discriminare mai redusă la mijlocul scalei (unde se îngrămădesc laolaltă numeroase rezultate); o repartiție uniformă a scorurilor asigură o discriminare la fel de bună de-a lungul întregii scale. O repartiție asimetrică (în care un număr disproporționat de mare de rezultate se acumulează la un capăt al scalei) este extrem de discriminativă la capătul unde găsim puține scoruri.

Se deduce, deci, că un test discriminează cu atât mai bine în cadrul întregului lot de subiecți, cu cât permite un număr mai mare de valori de scor și cu cât repartiția acestora este mai uniformă.

Pentru un test alcătuit din itemi dihotomici, cotați cu 0 și 1, al cărui scor se obține prin însumarea cotelor itemilor, o măsură a puterii de discriminare bazată pe forma repartiției scorurilor obținute la test, o reprezintă coeficientul lui Ferguson (Kline, 1993). Acesta se calculează cu ajutorul formulei:

$$\delta = \frac{n+1}{nN^2} (N^2 - \sum_{i=0}^n f_i^2) \quad (10.2.1)$$

unde:

- δ = coeficientul lui Ferguson;
- n = numărul itemilor testului;
- N = numărul de subiecți cărora li s-a administrat testul;
- f_i = frecvența scorului i , $i=0,1,\dots,n$.

Coeficientul lui Ferguson ia valori între 0 (atunci când apare un singur scor) și 1 (în cazul repartiției uniforme). Repartiția legii normale furnizează un coeficient egal cu 0.93. Se consideră că un test are o capacitate de discriminare bună dacă δ ia o valoare mai mare decât 0.9.

Există o legătură puternică între dificultatea unui test și sensibilitatea sa. Atunci când testul este foarte greu sau foarte ușor pentru un grup de persoane, rezultatele se concentrează spre una dintre extremitățile intervalului

de valori posibile pentru scoruri și, deci, nu sunt evidențiate deosebiri între indivizi.

Se impune o observație: forma repartiției de frecvențe a scorurilor este dependentă atât de forma repartiției de frecvențe a valorilor variabilei măsurate, în grupul testat, cât și de test (de caracteristicile psihometrice ale itemilor acestuia).

Exemplu (Albu, Pitariu, 1992a).

S-a considerat un test alcătuit din patru itemi, ale căror funcții caracteristice au expresia analitică:

$$P(y) = c + \frac{1-c}{1+e^{-1.7a(y-b)}}$$

S-a presupus că pentru o grupă de subiecți variabila măsurată de test are o repartiție simetrică.

Cu ajutorul calculatorului s-a determinat repartiția de frecvențe a scorurilor testului pentru diverse alegeri ale parametrilor a , b și c . În figura 10.2.1 sunt prezentate două dintre repartițiile de frecvențe obținute. Se constată că între forma repartiției variabilei măsurate în eșantionul de subiecți și forma repartiției scorurilor realizate pot exista deosebiri importante.

Atunci când repartiția de frecvențe a scorurilor obținute are o asimetrie puternică, dificultatea și sensibilitatea testului pot fi analizate identificând scorurile care apar mai frecvent. În celelalte cazuri, însă, pentru a putea aprecia corect aceste proprietăți ale testului este necesar să se calculeze principalii indicatori statistici ai scorurilor și să se compare, folosind teste statistice, valorile lor cu cele înregistrate în manualul testului sau obținute în alte ocazii.

Media rezultatelor aduce informații despre nivelul general al variabilei măsurate de test. Dacă media grupului examinat este semnificativ mai scăzută (respectiv, mai ridicată) decât cea a populației care a servit la etalonarea testului se deduce că testul este prea dificil (respectiv, prea ușor) pentru lotul respectiv și interpretarea scorului unei persoane prin raportare la etalonul din manual ar putea conduce la concluzii greșite.

Abaterea standard servește la aprecierea omogenității grupului examinat. O valoare mică a abaterii standard este caracteristică unui lot de subiecți foarte asemănători între ei în privința variabilei măsurate de test. În acest caz, testul nu este sensibil, nu realizează o diferențiere bună între persoane. Valorile mari ale abaterii standard se întâlnesc în loturile neomogene, cu deosebiri mari între cei testați. Dar atunci media scorurilor este mai puțin

ilustrativă pentru nivelul lotului, întrucât rezultatele obținute sunt împrăștiate într-un interval destul de mare în jurul valorii medii.

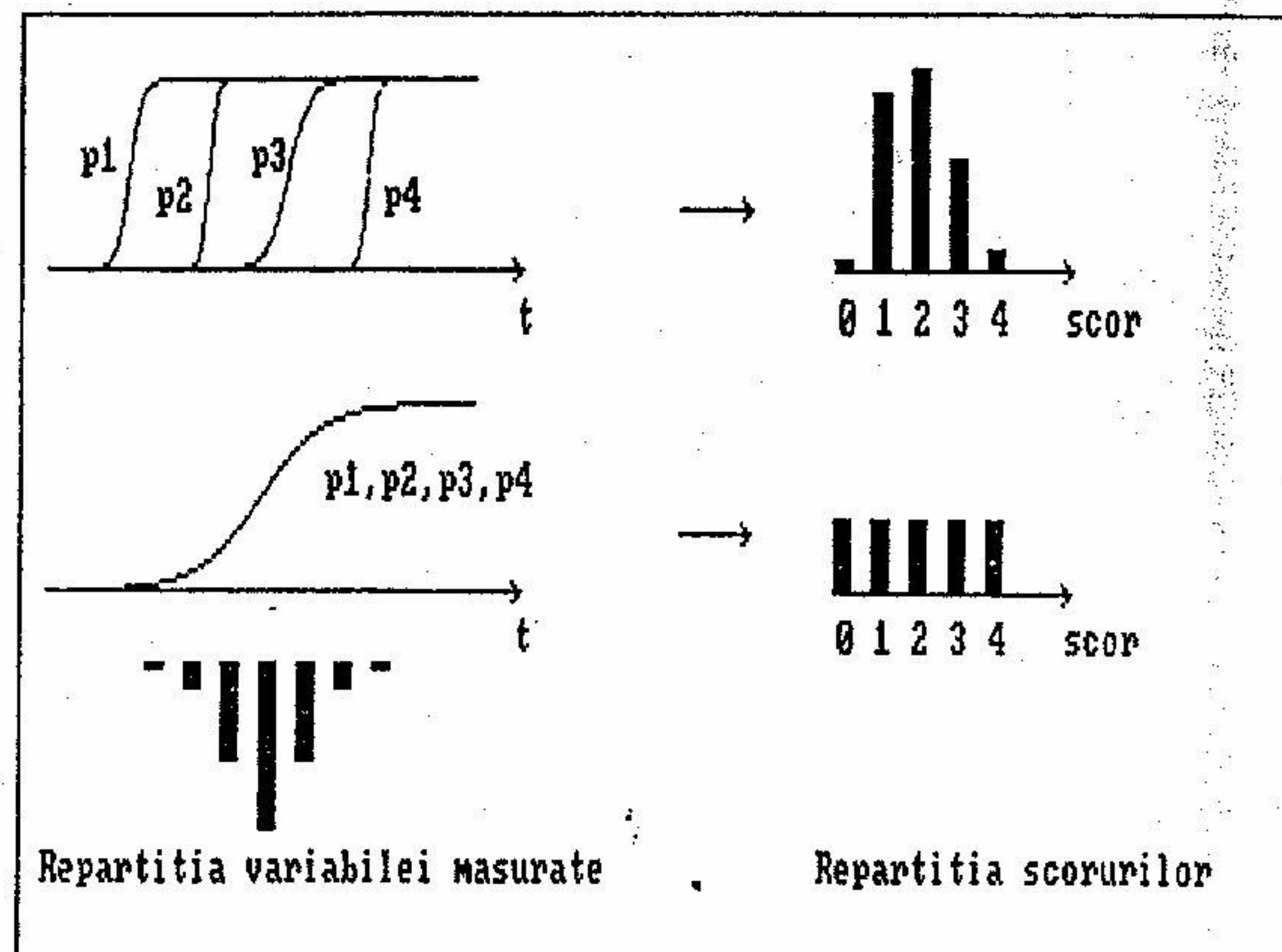


Figura 10.2.1

Comparatie între repartiția valorilor variabilei măsurate de test și repartiția scorurilor testului

Trebuie observat că nici un test nu poate fi atât de sensibil încât totdeauna două persoane deosebite în privința variabilei măsurate de test să obțină scoruri diferite. Aceasta, în primul rând, deoarece trăsăturile psihice, aptitudinile, cunoștințele și celelalte caracteristici măsurate de teste sunt variabile continue, în timp ce scorurile testelor sunt variabile discrete, având doar un număr finit de valori. Prin urmare, nici un test nu poate stabili o corespondență biunivocă între variabila măsurată și scorurile pe care le produce. Oricât de deosebiți între ei ar fi indivizii dintr-un lot examinat, dacă numărul lor este mai mare decât numărul de valori de scor pe care le poate furniza testul, vor exista persoane cu același scor, deci între care testul nu diferențiază.

Pe de altă parte, scorul realizat de o persoană la un test este afectat de diverse erori (erori cauzate de starea de oboseală a subiectului, de condițiile în care s-a făcut examinarea și cotașarea sau chiar erori de construcție a testului etc.). Din acest motiv, la fiecare nivel al variabilei măsurate, t , sunt posibile, cu probabilități diferite în general, mai multe valori de scor.

Dacă testul este bine construit, atunci scorul cel mai probabil variază nedescrescător față de variabila măsurată. Testul realizează în acest caz o descompunere a axei reale - pe care variază variabila măsurată - în intervale disjuncte, pe fiecare interval scorul cel mai probabil fiind același. Numărul acestor intervale nu poate fi mai mare decât numărul de valori de scor pe care le poate furniza testul. Mai mulți indivizi care au valori ale variabilei măsurate de test diferite între ele dar situate în același interval au șanse mari să obțină același scor, egal cu scorul cel mai probabil. Deci, testul nu va discrimina între ei.

Din cele arătate până acum se deduce că puterea de discriminare a unui test este legată de:

- numărul intervalelor pe care scorul cel mai probabil este același;
- lungimile acestor intervale;
- pozițiile intervalelor pe axa reală.

Un test va discrimina bine între indivizii care au valori ale variabilei măsurate de test plasate în acea zonă a axei reale în care există un număr mare de intervale, de lungimi mici.

10.3. INFLUENȚA CARACTERISTICILOR ITEMILOR ASUPRA DIFICULTĂȚII ȘI A SENSIBILITĂȚII TESTULUI

În continuare vom arăta, prin câteva exemple, cum depind dificultatea și sensibilitatea unui test de caracteristicile itemilor componenți.

Ne referim doar la teste alcătuite din itemi care măsoară aceeași variabilă, dihotomici, cotați cu 1 (pentru valori mari ale variabilei) și 0 (în cazul valorilor mici ale variabilei).

Pentru un asemenea item, indicele de dificultate este dat de frecvența relativă a persoanelor care au obținut scorul 1 în mulțimea subiecților care au răspuns la item. El arată cât de dificil este itemul pentru lotul de persoane testat.

În toate exemplele analizate itemii sunt valizi, în sensul că pentru fiecare item scorul 0 este obținut de persoane care la variabila măsurată de test au valori mai mici decât persoanele care obțin la item scorul 1.

Să considerăm 10 persoane s_1, s_2, \dots, s_{10} care au răspuns la 6 itemi, i_1, i_2, \dots, i_6 . Scorul testului format din acești itemi se obține prin însumarea scorurilor itemilor.

Notăm cu c_1, c_2, \dots, c_{10} valorile variabilei măsurate de test pentru cele 10 persoane și presupunem că între ele există relațiile:

$$c_1 < c_2 < \dots < c_{10} \quad (10.3.1)$$

Exemplul 10.3.1

Scorurile obținute de subiecți la itemi și la testul compus din ei sunt notate în Tabelul 10.3.1.

Se constată că dacă itemii componenți au aceeași dificultate și sunt intercorelați, testul nu discriminează bine între subiecți.

Alegerea unor itemi valizi cu aceeași valoare a indicelui de dificultate micșorează numărul scorurilor distincte obținute la test. Acest lucru se observă și din următorul exemplu.

Tabelul 10.3.1

Scorurile itemilor i_1, i_2, \dots, i_6 pentru subiecții s_1, s_2, \dots, s_{10}

Item	Persoana										Dificultate
	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	
i_1	0	0	0	0	0	1	1	1	1	1	0.5
i_2	0	0	0	0	0	1	1	1	1	1	0.5
i_3	0	0	0	0	0	1	1	1	1	1	0.5
i_4	0	0	0	0	0	1	1	1	1	1	0.5
i_5	0	0	0	0	0	1	1	1	1	1	0.5
i_6	0	0	0	0	0	1	1	1	1	1	0.5
Scor total	0	0	0	0	0	6	6	6	6	6	

Exemplul 10.3.2

Scorurile obținute de subiecți la itemi și la testul compus din ei sunt notate în Tabelul 10.3.2.

Tabelul 10.3.2

Scorurile itemilor i_1, i_2, \dots, i_6 pentru subiecții s_1, s_2, \dots, s_{10}

Item	Persoana										Dificultate
	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	
i_1	0	0	0	0	0	0	1	1	1	1	0.4
i_2	0	0	0	0	0	0	1	1	1	1	0.4
i_3	0	0	0	0	0	0	1	1	1	1	0.4
i_4	0	0	0	0	1	1	1	1	1	1	0.6
i_5	0	0	0	0	1	1	1	1	1	1	0.6
i_6	0	0	0	0	1	1	1	1	1	1	0.6
Scor total	0	0	0	0	3	3	6	6	6	6	

Din cele 7 valori posibile de scor la test s-au obținut doar 3. Scorurile testului nu discriminează între subiecții cu valori mari ale variabilei măsurate (care obțin, toți, scorul maxim 6) și nici între cei cu valori mici ale acestei variabile (care obțin, toți, scorul minim, 0). Cauza o constituie faptul că toți itemii componenți corelează între ei și sunt de dificultate medie.

Exemplul 10.3.3

Scorurile obținute de subiecți la itemi și la testul compus din ei sunt notate în Tabelul 10.3.3.

Din cele 7 valori posibile de scor la test s-au obținut 5: 0, 1, 3, 5 și 6. Scorurile testului sunt repartizate aproximativ echidistant în intervalul scorurilor posibile, [0,6], datorită faptului că și dificultățile itemilor care compun testul sunt aproximativ echidistante în intervalul [0,1]. Testul format este de dificultate medie.

Tabelul 10.3.3

Scorurile itemilor i_1, i_2, \dots, i_6 pentru subiecții s_1, s_2, \dots, s_{10}

Item	Persoana										Dificultate
	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	
i_1	0	0	0	0	0	0	0	0	1	1	0.2
i_2	0	0	0	0	0	0	1	1	1	1	0.4
i_3	0	0	0	0	0	0	1	1	1	1	0.4
i_4	0	0	0	0	1	1	1	1	1	1	0.6
i_5	0	0	0	0	1	1	1	1	1	1	0.6
i_6	0	0	1	1	1	1	1	1	1	1	0.8
Scor total	0	0	1	1	3	3	5	5	6	6	

Cu totul alta este situația în următorul exemplu.

Exemplul 10.3.4

Tabelul 10.3.4

Scorurile itemilor i_1, i_2, \dots, i_6 pentru subiecții s_1, s_2, \dots, s_{10}

Item	Persoana										Dificultate
	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	
i_1	0	0	0	0	0	0	0	1	1	1	0.3
i_2	0	0	0	0	0	1	1	1	1	1	0.5
i_3	0	0	0	0	1	1	1	1	1	1	0.6
i_4	0	0	0	1	1	1	1	1	1	1	0.7
i_5	0	0	0	1	1	1	1	1	1	1	0.7
i_6	0	0	0	1	1	1	1	1	1	1	0.7
Scor total	0	0	0	3	4	5	5	6	6	6	

Scorurile obținute de subiecți la itemi și la testul compus din ei sunt notate în Tabelul 10.3.4. Și în acest caz s-au obținut tot 5 scoruri la test, dar acestea sunt mai apropiate de scorul maxim posibil decât în exemplul precedent, testul este mai ușor. Aceasta, din cauza faptului că itemii care compun testul au indici de dificultate mai mare decât în exemplul precedent.

Pentru a obține cât mai multe valori de scor se aleg itemi care să aibă indici de dificultate diferiți și repartizați cât mai uniform pe intervalul $[0,1]$, ca în exemplul următor.

Exemplul 10.3.5

Scorurile obținute de subiecți la itemi și la testul compus din ei sunt notate în Tabelul 10.3.5.

Tabelul 10.3.5

Scorurile itemilor i_1, i_2, \dots, i_6 pentru subiecții s_1, s_2, \dots, s_{10}

Item	Persoana										Dificultate
	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	
i_1	0	0	0	0	0	0	0	0	0	1	0.1
i_2	0	0	0	0	0	0	0	0	1	1	0.2
i_3	0	0	0	0	0	0	1	1	1	1	0.4
i_4	0	0	0	0	1	1	1	1	1	1	0.6
i_5	0	0	1	1	1	1	1	1	1	1	0.8
i_6	0	1	1	1	1	1	1	1	1	1	0.9
Scor total	0	1	2	2	3	3	4	4	5	6	

De data aceasta apar toate scorurile posibile ale testului, iar repartitia lor de frecvențe este cuprinsă între repartitia uniformă și cea normală.

CAPITOLUL 11 INFORMAȚII DESPRE TESTELE PSIHOLOGICE

11.1. MANUALELE TESTELOR PSIHOLOGICE

În mod obișnuit, orice test psihologic care este pus în circulație, pentru a fi utilizat în practică, este însoțit de două manuale:

- un *manual tehnic*, care conține informații despre felul în care a fost construit testul, rezultatele analizei de itemi, valorile coeficienților de fidelitate și concluziile studiilor de validitate;

- un *manual de utilizare*, care cuprinde instrucțiunile pentru administrarea testului și indicații pentru interpretarea rezultatelor.

Standards for Educational and Psychological Tests din 1985 recomandă ca, pe lângă furnizarea acestor informații, manualele testelor psihologice să îndeplinească și următoarele sarcini (Gregory, 1992):

a. să descrie motivele construirii testului și să prezinte recomandări pentru utilizarea lui;

b. să atragă atenția asupra eventualelor riscuri de folosire greșită a testului;

c. să citeze studii reprezentative privind utilizarea testului în situații obișnuite sau specifice;

d. să prezinte condițiile (de studii, de experiență etc.) pe care trebuie să le îndeplinească persoana care administrează testul și cea care interpretează rezultatele acestuia;

e. să descrie relațiile constatate între scorurile testului și diverse criterii;

f. să prezinte posibilitățile de interpretare a rezultatelor testului cu ajutorul calculatorului.

11.2. INFORMAȚII NECESARE PENTRU ALEGEREA UNUI TEST PSIHOLOGIC

Atunci când sunt disponibile mai multe teste care servesc aceluiași scop, alegerea celui mai potrivit dintre ele impune luarea în considerare a unor criterii de ordin tehnic, practic și economic. Faptul că un test este un instrument de măsură sau de predicție bun, adecvat lotului care urmează să fie examinat și

accesibil ca preț, nu surmontează dificultățile legate de administrarea și de cotarea sa, ori de interpretarea rezultatelor.

Principalele aspecte care pot influența alegerea unui test psihologic sunt cuprinse în lista următoare (adaptată după Gronlund, 1967; Thorndike, Hagen, 1961):

A. Date pentru identificarea testului:

- denumirea testului;
- autorii testului (numele și, eventual, autoritatea lor științifică);
- versiunea testului;
- data punerii în circulație;
- denumirea firmei care difuzează testul.

B. Informații generale despre test:

- funcția testului: ce măsoară și cum este definit constructul măsurat sau ce performanțe prezice;
- caracteristicile persoanelor care pot fi examinate cu acest test: zona geografică, vârsta, nivelul de școlarizare etc.;
- timpul necesar pentru administrarea testului;
- modul de administrare: creion-hârtie, cu aparate, cu ajutorul calculatorului etc.;
- modul de calcul al scorurilor: manual, cu calculatorul etc.;
- cum se interpretează scorurile testului: obiectiv sau subiectiv;
- dacă există forme paralele ale testului, disponibile;
- costul achiziționării testului;
- costul administrării testului pentru o persoană.

C. Caracteristici tehnice:

- calitatea itemilor testului: dacă folosesc o terminologie cunoscută celor care vor fi examinați cu acest test, dacă sunt lipsiți de ambiguitate, dacă în itemii cu alegere multiplă răspunsurile greșite propuse nu sunt banale;
- validitatea: modul în care a fost studiată validitatea și constatările făcute;
- fidelitatea: ce coeficienți au fost calculați, în ce condiții și ce rezultate s-au obținut;
- norme disponibile: ce tipuri de norme se utilizează, ce populații au fost folosite pentru etalonare, dacă există norme potrivite pentru populația căreia urmează să îi fie administrat testul.

D. Caracteristici practice:

- dificultăți la administrare: dacă sunt necesare aparate sau condiții speciale;

- dificultăți la calculul scorurilor: dacă se aplică formule complicate și nu există program pentru calculul automatizat al scorurilor sau dacă este un test subiectiv, care necesită o interpretare calitativă a răspunsurilor subiecților;

- dificultăți la interpretarea scorurilor: dacă persoana care urmează să interpreteze scorurile trebuie să îndeplinească anumite condiții de studii sau de experiență;

- calitatea manualului testului: dacă există suficiente informații pentru utilizarea testului și dacă acestea sunt clare.

E. Alte considerații:

- părerile unor persoane care au utilizat deja testul.

Majoritatea acestor informații se găsesc în manualele testului. Cele referitoare la existența formelor paralele ale testului sau a programelor pentru administrarea, cotarea și/sau interpretarea rezultatelor testului se pot obține din reviste de specialitate. Informațiile privind distribuitorii testului, costul de achiziție și cel de utilizare (de administrare și/sau de cotare) sunt specifice fiecărei țări.

Foarte multe reviste publică articole despre validitatea și fidelitatea testelor. Cel mai bun mijloc pentru identificarea lucrărilor referitoare la un anumit test îl constituie revista de rezumate *Psychological Abstracts*.

Se pot obține multe informații despre testele psihologice din următoarele reviste:

- *Advances in Personality Assessment*;
- *The Journal of Psychoeducational Assessment*;
- *Psychology in the School*;
- *Educational and Psychological Measurement*;
- *Psychological Assessment: A Journal of Clinical and Consulting Psychology*;
- *The Journal of Clinical Psychology*;
- *The Journal of Personality Assessment*.

În *Revista de Psihologie* și în *Revue Roumaine de Psychologie* sunt publicate articole despre studiile efectuate în România asupra unor teste psihologice originale sau traduse.

CAPITOLUL 12 CONSTRUIREA TESTELOR PSIHOLOGICE

12.1. ALGORITMUL GENERAL DE CONSTRUIRE A TESTELOR

12.1.1. Prezentarea pașilor algoritmului de construire a unui test

În general, construirea unui test psihologic sau a unei scale dintr-un test se desfășoară în mai mulți pași, nescvențial, cu reveniri repetate până la realizarea calităților dorite (fidelitate, validitate, putere de discriminare, dificultate). Algoritmul general de construire a unui test este redat în figura 12.1.1.1.

Conținutul tuturor pașilor algoritmului este determinat de direcțiile stabilite la primul pas, în special de scopul - teoretic și practic - căruia trebuie să îi servească elaborarea testului. În funcție de acesta se formulează itemii, se alege metoda de construire a testului și se alcătuiește o primă variantă a testului.

Gruparea de itemi formată se administrează unui eșantion de persoane extras din populația căreia îi este destinat testul. Se analizează răspunsurile obținute la fiecare item, se identifică itemii necorespunzători (de exemplu, cei la care majoritatea persoanelor au răspuns identic sau, în cazul itemilor cu alegere multiplă, cei la care un răspuns greșit s-a dovedit a fi mai atractiv pentru subiecți decât răspunsul corect) și se caută cauza, pentru a vedea dacă este legată de item, de eșantionul de subiecți sau de modul de administrare a itemilor. Analiza cauzelor erorilor poate scoate la iveală deficiențe ale unor itemi care în urma analizei statistice a rezultatelor păreau a fi bine construiți (de exemplu, la un item s-a răspuns greșit pentru că un alt item conține informații eronate, care i-au derutat pe subiecți). Itemii fără greșeli se introduc în partea a doua a analizei de itemi, care aplică procedee specifice metodei alese pentru construirea testului (de exemplu, studiul validității itemilor sau calculul puterii de discriminare a acestora). Este posibil să fie puși în evidență alți itemi necorespunzători. Din itemii fără deficiențe se alcătuiește un test, folosind un algoritm specific metodei de construcție alese la primul pas al algoritmului.

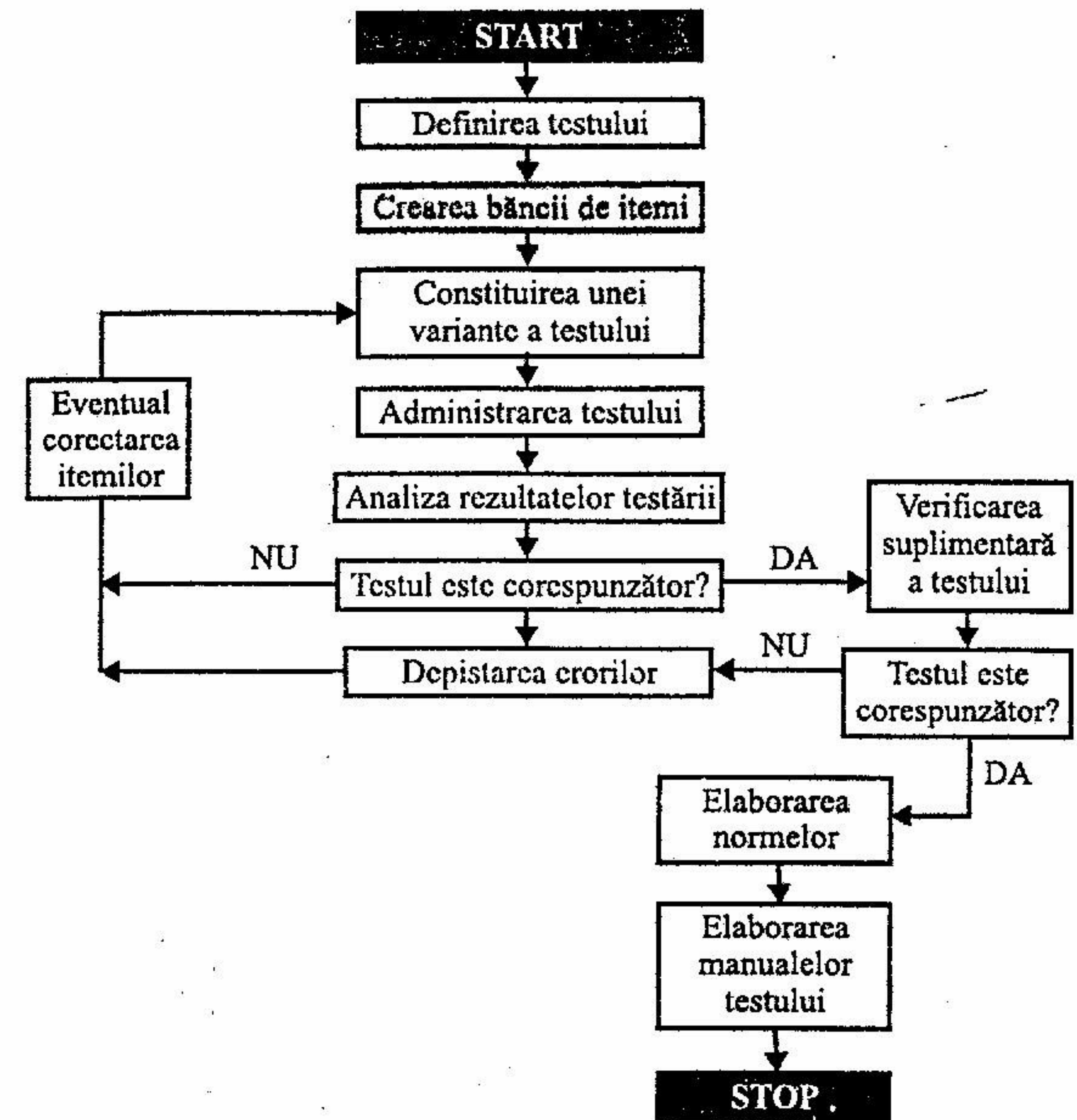


Figura 12.1.1.1

Algoritmul general de construire a unui test psihologic

Se studiază caracteristicile acestui test (fidelitatea, validitatea, puterea de discriminare, dificultatea) și, dacă sunt considerate acceptabile, se administrează testul unui nou eșantion de persoane, similar primului, pentru a verifica dacă se păstrează calitățile.

Atunci când pentru unul din eșantioane testul apare ca fiind inacceptabil, se alcătuiește o nouă variantă a testului și se reia procedura descrisă mai sus.

Dacă se mențin calitățile testului și la verificarea suplimentară, se determină normele testului și se stabilesc regulile de interpretare a scorurilor. După ce se redactează manualele, testul poate fi pus în circulație.

12.1.2. Observații asupra metodelor de construire a testelor

În majoritatea lucrărilor de teoria testelor psihologice, metodele de construire a testelor sunt împărțite în următoarele trei categorii (Burisch, 1986; Van Der Maesen, Hofstee, 1989):

A. Metodele deductive, denumite și raționale sau bazate pe simțul comun, sunt cele în care construirea itemilor și selectarea celor care vor compune testul se bazează exclusiv pe părerile unor persoane (experți sau nu). Mai întâi sunt alese constructele care vor fi măsurate de teste, apoi sunt elaborați itemii, care realizează o definiție implicită sau explicită a constructelor, singurul ghid în această operație fiind intuiția autorilor itemilor. Din această categorie fac parte:

- metoda prototipului;
- metoda BARS;
- metoda designului cu fațete;
- metoda intervalelor aparent egale.

B. Metodele inductive sau interne pornesc de la o colecție mare, eterogenă, de itemi (de obicei extrași din teste aflate în circulație), considerată a fi reprezentativă pentru constructele care trebuie măsurate. Prin procedee statistice, cum sunt analiza factorială și analiza de clusteri, se urmărește să se depisteze grupările de itemi care corespund constructelor respective. Din această categorie de metode fac parte:

- metoda consistenței interne;
- metoda teoretică;
- metoda analizei criteriului intern.

C. Metodele externe sau empirice necesită precizarea, de la început, a relațiilor care vor trebui să existe între scorurile testului care va fi construit și diverse criterii externe. Se pornește de la o colecție mare, eterogenă, de itemi, se compară scorurile acestora cu criteriile și se rețin cei care au comportarea pe care trebuie să o aibă întregul test. Din această categorie de metode fac parte:

- metoda criteriului extern;
- metoda analizei constructului.

După cum se observă, metodele deductive nu necesită nici experimentarea itemilor, nici efectuarea unor analize statistice, în timp ce metodele inductive și cele externe se bazează pe prelucrarea statistică a rezultatelor experimentării itemilor, acordând o pondere redusă modului de construire a itemilor.

În multe situații este posibilă - și chiar recomandabilă - construirea testelor cu ajutorul mai multor metode: o metodă deductivă va servi la formarea băncii de itemi, iar printr-o metodă internă sau/și una externă se vor selecta itemii care vor alcătui testul.

Din acest motiv, în cadrul algoritmului de construire a testelor metodele deductive vor fi prezentate la pasul "Crearea băncii de itemi", iar metodele interne și cele externe, la pasul "Analiza rezultatelor testului".

12.2. DEFINIREA TESTULUI

După ce se ia decizia de a construi un test, se precizează în detaliu:

- a. funcția pe care o va îndeplini (măsurare sau predicție);
- b. populația căreia îi este destinat;
- c. condițiile în care va fi administrat: când, unde, cum, în cât timp;
- d. modul de interpretare a rezultatelor (prin raportare la obiective fixate a priori sau prin raportare la norme);
- e. motivele care au impus crearea lui.

Aceste aspecte determină toate acțiunile legate de elaborarea și experimentarea sa.

Pornind de la funcția pe care o va realiza testul se decide care dintre caracteristicile sale psihometrice este mai importantă (fidelitatea sau validitatea), întrucât se știe că un test nu poate avea în același timp și o validitate relativă la criteriu foarte bună, și o consistență internă ridicată. De asemenea, un test poate fi un instrument de măsură excelent, fără a avea vreo valoare predictivă.

Modul de interpretare a rezultatelor unui subiect este dependent de scopul căruia îi va servi testul. Atunci când testul verifică dacă anumite deprinderi sau cunoștințe ale unei persoane depășesc un prag minim necesar, scorul individual va fi raportat la bareme fixate a priori. Dacă, însă, testul prezice comportamentul unei persoane în situații specifice sau determină cât de eficient este individul într-un anumit domeniu, comparativ cu alte persoane din populația căreia îi aparține subiectul, atunci interpretarea scorului individual se face prin raportare la norme.

Tot la acest prim pas al algoritmului de construire a testului se decide:

- a. ce fel de scală de măsură va utiliza testul (nominală, ordinală sau de interval);

- b. ce tipuri de itemi se vor include în test (în funcție de ceea ce va măsura sau va prezice testul și de faptul că pentru cotaarea răspunsurilor se va utiliza sau nu calculatorul);
- c. cum vor fi ordonați itemii (după dificultate sau aleator);
- d. ce lungime va avea testul (în funcție de caracteristicile populației căreia i se va administra testul și de timpul de administrare);
- e. cum va fi administrat testul: cu sau fără limită de timp

A. Atunci când noul test trebuie să măsoare un construct, este de dorit ca la baza sa să stea o teorie structurală asupra personalității, care să reflecte și importanța pentru viață a diferențelor individuale în privința acestuia (Angleitner, John, Löhr, 1986). Definirea constructului care va fi măsurat de test va specifica formele observabile relevante de manifestare a constructului respectiv (comportamente, interese, dorințe) și, în plus, va prezenta relațiile care, conform teoriei, se așteaptă să existe între acesta și alte constructe (atât relații convergente, cât și discriminante). Astfel de relații se pot deduce ușor dacă pentru constructul măsurat de test se construiește o rețea nomologică.

Exemplu (Gregory, 1992).

A.S. Kaufman și N.L. Kaufman, când și-au propus să construiască un test pentru măsurarea inteligenței, au formulat condițiile impuse acestuia astfel:

- Să măsoare inteligența, având la bază o teorie puternică și rezultate ale cercetărilor.
- Să facă distincție între cunoștințele dobândite și abilitatea de a rezolva probleme cu care subiectul nu este familiarizat.
- Să furnizeze scoruri care să poată fi utilizate pentru intervenții educaționale.
- Să includă sarcini noi.
- Să fie ușor de administrat.
- Să fie cotate obiectiv.
- Să fie sensibil la diverse caracteristici ale preșcolarilor, ale grupurilor minoritare și ale copiilor cu necesități speciale.

B. Un test care urmează să fie folosit în selecția profesională trebuie să poată prezice corect performanța persoanelor într-o anumită muncă. Pentru asemenea teste este necesar să se realizeze, de la început:

- a. analiza muncii respective: să se obțină informații despre sarcinile de muncă și despre cerințele postului;
- b. descrierea muncii: să se prezinte în detaliu, pentru munca respectivă, sarcinile, responsabilitățile și condițiile de lucru;

c. specificarea muncii: să se stabilească ce caracteristici fizice și psihice sunt necesare desfășurării muncii respective, alături de care se vor evidenția (Kline, 1993):

- deprinderile și cunoștințele care nu sunt absolut necesare, deși simțul comun ar sugera că ele sunt importante pentru îndeplinirea muncii (de exemplu, cunoștințele de anatomie nu le sunt absolut necesare psihiatrilor, ele intervenind extrem de rar în activitatea lor zilnică);
- caracteristicile necesare datorită contextului în care se desfășoară munca respectivă. (De exemplu, un flautist profesionist trebuie să fie capabil să lucreze într-o orchestră mare, să fie familiarizat cu repertoriul ultimelor trei secole, să poată să cânte în concerte după un număr mic de repetiții, să poată suporta efortul unor călătorii lungi și să fie dispus să lipsească mult de acasă. Aceste cerințe sunt diferite de cele pentru un flautist de jazz sau pentru un flautist într-o formație muzicală mică.)

C. În cazul testelor de cunoștințe, definirea testului constă în descrierea domeniului de conținut. Adesea, se construiește o tabelă de specificații, care indică repartitia itemilor testului în funcție de obiectivele comportamentale urmărite de test și de temele a căror cunoaștere se verifică. O asemenea tabelă este prezentată în tabelul 12.2.1.

Tabelă de specificații servește ca bază pentru a decide ce fel de itemi se vor introduce în test, întrucât pentru fiecare celulă a tabelului sunt mai potrivite anumite tipuri de itemi (de exemplu, capacitatea de explicare a unor fapte poate fi verificată mai bine prin itemi de tip eseu decât prin itemi cu răspunsuri închise).

Apariția unui test nou trebuie să răspundă unei necesități reale. Se recurge la procedura - costisitoare - de creare a unui test atunci când nu există un test care să îndeplinească scopul urmărit (măsurarea unui anumit construct, verificarea cunoștințelor dintr-un anumit domeniu sau predicția unui anumit comportament), dacă testele disponibile sunt dificil de administrat, de cotate sau/și de interpretat ori sunt prea scumpe. Uneori se construiesc forme paralele ale unor teste existente, fie pentru a nu administra repetat același test, fie pentru a se evita transmiterea itemilor de la un eșantion de subiecți la altul și a putea păstra secret conținutul testului (de exemplu, în cazul testelor de cunoștințe utilizate la diverse concursuri).

Tabelul 12.2.1

Tabela de specificații pentru verificarea cunoștințelor la tema
"Teoria testelor psihologice"

Obiectivul comportamental	Subiectul		
	Validitatea testelor	Fidelitatea testelor	Analiza de itemi
Cunoașterea terminologiei	Validitatea relativă la: -construct; -conținut; -criteriu. (3 itemi)	Consistența internă. Stabilitatea în timp. Forme paralele ale testelor. (3 itemi)	Dificultatea unui item. Validitatea unui item. (2 itemi)
Cunoașterea faptelor specifice	Metoda de studiere a validității: -predictive; -concurente. (2 itemi)	Metodele: -test-retest; -formelor paralele; -analiza consistenței interne. (3 itemi)	Procedura de calcul pentru: -indicele de dificultate; -puterea de discriminare; -coeficientul de validitate. (3 itemi)
Înțelegere	Explicarea legăturii dintre validitate și fidelitate. Explicarea efectelor restrângerii intervalului de valori asupra coeficientului de validitate. (2 itemi)	Explicarea utilizării coeficientului de fidelitate pentru interpretarea scorurilor la test. (1 item)	Explicarea deosebirii dintre puterea de discriminare și validitatea itemilor. (1 item)
Aplicare	Calculul unui coeficient de validitate. (1 item)	Aplicarea formulei lui Spearman-Brown. Determinarea numărului de itemi care trebuie adăugați unui test pentru dublarea valorii coeficientului de fidelitate. (2 itemi)	Calculul indicelui de dificultate și al puterii de discriminare. (2 itemi)
TOTAL	8 itemi	9 itemi	8 itemi

12.3. CREAREA BĂNCII DE ITEMI

Itemii care se introduc în banca de itemi pot să provină din mai multe surse:

a. sunt formulați de psihologi, pe baza unei teorii (dacă testul măsoară un construct psihologic);

b. sunt elaborați de specialiști din domeniul în care va fi aplicat testul (dacă acesta evaluează deprinderi ori cunoștințe însușite în urma unui program de instruire sau dacă se utilizează în selecția profesională);

c. sunt creați de diverse persoane (nu neapărat experți), ca ilustrări ale comportamentelor pe care acestea le consideră caracteristice unui anumit construct;

d. sunt extrași din teste existente, care au fost administrate unui lot de persoane asemănător celui căruiia îi este destinat noul test.

Ultimul procedeu se aplică de obicei atunci când s-a constatat că itemii respectivi conțin un factor comun care corespunde constructului pe care îl va măsura testul sau când aceștia îndeplinesc cerințele care decurg din scopul în care va fi utilizat testul. De exemplu:

- au cu alte variabile legături conforme teoriei care stă la baza constructului

sau

- se încadrează în domeniul de conținut al testului

sau

- corelează semnificativ cu variabila criteriu

sau

- discriminează între două grupe contrastante de persoane.

Înainte de introducerea itemilor în bancă se stabilește modul în care vor fi cotați. Uneori, se apreciază dificultatea acestora, atunci când ea este importantă pentru alegerea cotelor care vor fi acordate răspunsurilor sau dacă în testul care va fi format itemii vor fi aranjați în ordinea crescătoare a dificultății.

Se recomandă ca numărul itemilor introduși în bancă să fie cel puțin cu 20% mai mare decât lungimea testului (Aiken, 1994).

După ce itemii au fost scriși, este bine ca ei să fie revăzuți și editați de o persoană care nu a participat la elaborarea lor. Ea poate sesiza eventualele erori și poate face propuneri pentru îmbunătățirea itemilor.

Există firme specializate în crearea de teste, cum este, de exemplu, Educational Testing Service. Acestea folosesc pentru scrierea itemilor persoane care au cunoștințe din domeniul la care se va referi testul și, în plus, au deprinderea de a crea itemi.

Cei care doresc să elaboreze itemi de test pot utiliza drept model itemii din testele standardizate comercializate.

Prezentăm în continuare câteva metode de creare a băncii de itemi, din cadrul metodelor deductive de construire a testelor. În cazul metodelor interne și externe de construire a testelor, alcătuirea băncii de itemi este mai simplă: itemii -noi sau extrași din alte teste- nu mai sunt verificați înainte de a fi introduși în bancă, întrucât după administrarea lor unui eșantion de subiecți rezultatele obținute vor fi prelucrate statistic.

Metoda prototipului (Oosterveld, Vorst, 1995) se utilizează pentru construirea chestionarelor de personalitate.

Mai întâi sunt generați itemii. În acest scop, persoanelor care participă la această acțiune li se cere să descrie comportamente tipice pentru indivizi care au valori foarte mari la constructul măsurat de test. De exemplu, pentru elaborarea unui test care să evalueze dominanța, se poate folosi instrucțiunea: "Gândiți-vă la trei bărbați (femei) care sunt cei mai dominanți dintre cunoștințele dumneavoastră. Descrieți cinci acțiuni sau comportamente ale acestora prin care puteți exemplifica de ce îi considerați dominanți."

Se analizează apoi mulțimea de descrieri obținute, se elimină redundanțele și se corectează greșelile de exprimare. Fiecare descriere rămasă va constitui un item.

Pasul următor constă din evaluarea itemilor de către mai multe persoane. Se acordă fiecărui item câte o cotă, de exemplu între 1 și 5, care să reflecte cât de tipic este comportamentul descris de item pentru cei care au valori mari la constructul măsurat de test.

Se calculează pentru fiecare item media cotelor primite de la toți evaluatorii și se rețin în bancă itemii care au mediile cele mai mari.

O variantă a acestei metode (Angleitner, John, Löhr, 1986) pornește de la definirea mai multor constructe, pe baza unui model structural al comportamentului interpersonal. Un grup numeros de subiecți elaborează, pentru fiecare construct, itemi care descriu manifestări comportamentale specifice constructului. Apoi, fiecare item este analizat de persoane dintr-un alt grup și este atribuit constructului pe care se consideră că îl ilustrează cel mai bine. Sunt reținuți în banca de itemi doar acei itemi care au fost alocați, cu frecvență mare, constructului pentru ilustrarea căruia au fost creați.

Metoda BARS (Behavioral Anchored Rating Scales) a fost aplicată la început pentru construirea unor scale de apreciere a comportamentului legat

de muncă. Fiecare scală corespundea unei dimensiuni a activității profesionale, iar ancorele erau exemple de niveluri de performanță. Mai nou, această metodă este utilizată și pentru elaborarea unor chestionare de autoevaluare comportamentală bazate pe ancore. Este cazul unor chestionare care măsoară constructe multidimensionale. În cadrul lor, fiecare scală se referă la câte o dimensiune.

Algoritmul de construire a unui asemenea chestionar este compus din următorii pași (Landy, Rastegary ș.a., 1991):

(1) Un grup de specialiști în domeniul pentru care se construiește chestionarul încearcă să identifice dimensiunile constructului. Fiecare dimensiune primește un nume, fără a fi, însă, definită.

(2) Lista cu denumiri este revăzută. Se elimină redundanțele. Unele denumiri sunt combinate între ele și reformulate, pentru a se obține categorii mai ușor de înțeles.

(3) Pentru fiecare "etichetă" a constructului se generează o definiție. Se contopesc dimensiunile pentru care, în urma analizei definițiilor, se constată existența unor suprapuneri.

(4) Se generează exemple pentru diverse niveluri de manifestare a dimensiunilor. De pildă, se pot fixa trei niveluri (înalt, mediu și scăzut) și i se cere fiecărei persoane din grupul de specialiști să formuleze câte trei ancore pentru fiecare nivel al fiecărei dimensiuni.

(5) Se elimină exemplele duble.

(6) Toate ancorele rămase sunt prezentate unui grup de evaluatori care au sarcina de a identifica dimensiunea la care se referă fiecare. Sunt reținute exemplele pentru care cel puțin 75% dintre persoane au recunoscut dimensiunea măsurată.

(7) Se elimină dimensiunile cărora li s-au repartizat doar puține ancore.

(8) Un alt grup de persoane atribuie valori dintr-o mulțime fixată (de exemplu, {1,2,...,7}) fiecărei ancore. Ponderea acordată trebuie să reflecte nivelul de dezvoltare al constructului măsurat de chestionar pentru care exemplul respectiv este potrivit.

(9) Pentru fiecare ancoră se calculează media și abaterea standard a cotelor primite. Se elimină exemplele care au abateri standard mari (de obicei, cele pentru care $\sigma > 1$); acestea au fost interpretate diferit de către evaluatori.

(10) Se ordonează ancorele fiecărei dimensiuni crescător după medie. Se formează scalele -câte una pentru fiecare dimensiune-reținând între 4 și 8 ancore, repartizate între nivelul cel mai scăzut și nivelul cel mai ridicat de manifestare a dimensiunii respective.

Asemănătoare cu metoda BARS este metoda ordonării comportamentelor de către experți (Gregory, 1992). Aceasta servește la crearea unor

teste în care itemii sunt așezați în ordinea crescătoare a dificultății și care au proprietatea că ori de câte ori un item este potrivit pentru o persoană, sunt potriviți într-o oarecare măsură și toți itemii care îl preced. Scorul unui subiect este dat de numărul de ordine al celui mai dificil item care i se potrivește.

Un exemplu de astfel de test îl reprezintă "Scala celui mai bun răspuns verbal" din Glasgow Coma Scale for Recording Depth of Coma, un test utilizat în spitale pentru evaluarea gradului de afectare a creierului în urma unui traumatism. Itemii componenți sunt:

- 1 - nici un răspuns;
- 2 - răspuns care nu poate fi înțeles;
- 3 - răspuns nepotrivit;
- 4 - răspuns confuz;
- 5 - răspuns la obiect.

Scorul scalei este stabilit de medic în urma observării pacientului.

Metoda designului cu fațete (Oosterveld, Vorst, 1995) are ca punct de pornire analiza constructului pe care îl va măsura testul și deducerea aspectelor mai importante ale acestuia, denumite "fațete". De exemplu, fațetele pot fi "situația" și "răspunsul".

Pentru fiecare fațetă se identifică elementele cele mai importante. Fațeta "răspuns", de pildă, poate avea ca elemente: "cognitiv", "psihologic", "comportamental" și "afectiv".

Se construiește apoi o matrice, cu atâtea dimensiuni câte fațete au fost alese, în care fiecare celulă corespunde unui element din fiecare fațetă. De exemplu, o celulă poate corespunde răspunsului "cognitiv" și situației "îmbolnăvire gravă". Această matrice este denumită "designul cu fațete".

Se scriu unul sau mai mulți itemi pentru fiecare celulă a matricii.

Metoda intervalelor aparent egale (Gregory, 1992; Kline, 1993) a fost propusă de L.L. Thurstone pentru construirea scalelor de evaluare a atitudinilor.

Se începe cu formarea unei colecții, cât mai bogate, de itemi dihotomici, care se referă, toți, la atitudinea față de un același obiect (persoană, instituție, situație sau noțiune abstractă). La fiecare item se poate răspunde doar prin "adevărat" sau "fals". Răspunsul ales de subiect este privit ca o reflectare a atitudinii sale, pozitive sau negative. De exemplu, pentru cercetarea atitudinii față de filmele muzicale se pot folosi itemii: "Un film muzical mă face optimist." și "Este o pierdere de vreme să urmăresc un film muzical".

Un grup de experți analizează fiecare item și apreciază cât de favorabilă este atitudinea studiată la persoanele care răspund cu "adevărat", acordând o

cotă cuprinsă între 1 (extrem de nefavorabilă) și 11 (extrem de favorabilă). Evaluatoarea li se cere să privească cele 11 cote posibile ca fiind echidistante.

Pentru fiecare item se calculează media și dispersia cotelor primite. Deoarece o dispersie mare indică o neconcordanță a părerilor, itemii cu dispersii mari sunt eliminați. Media cotelor va fi considerată drept scor corespunzător răspunsului "adevărat".

Fiecărei persoane examinate cu un astfel de test i se cere să indice itemii la care răspunsul său este "adevărat". Scorul total al testului pentru un subiect va fi egal cu media scorurilor itemilor indicați.

Scalele de acest tip se compun din 20-30 de itemi, aleși astfel încât scorurile corespunzătoare răspunsului "adevărat" să fie aproximativ echidistante în intervalul [1, 11].

O metodă asemănătoare celei descrise mai sus utilizează mediana în locul mediei, atât pentru determinarea scorurilor itemilor, cât și pentru calculul scorului unui subiect.

12.4. CONSTRUIREA UNEI VARIANTE A TESTULUI

Înainte de selectarea itemilor din banca de itemi, pentru alcătuirea unei variante a testului, se stabilește lungimea acestuia. Numărul de itemi care îl vor compune are influență atât asupra timpului de rezolvare a testului, cât și asupra calităților sale psihometrice (fidelitate, putere de discriminare).

Lungimea testului este aleasă în funcție de timpul disponibil pentru administrarea testului, care este fixat, în general, ținând cont de condițiile în care va fi utilizat testul (într-o oră de clasă, la un concurs de admitere într-o instituție de învățământ etc.).

Timpul de rezolvare a unui test este dependent nu numai de numărul itemilor componenți, ci și de:

- vârsta subiecților;
- deprinderea subiecților de a citi;
- lungimea și dificultatea itemilor;
- modul de formulare a răspunsurilor și lungimea acestora.

În practică s-a constatat că un test de dificultate medie, administrat elevilor din gimnaziu, necesită cam un minut pentru rezolvarea unui item cu alegere multiplă sau cu răspuns lacunar și aproximativ 30 de secunde pentru alegerea răspunsului la un item dihotomic. Prin urmare, într-o oră de clasă de 50 de minute se pot administra teste compuse din 100 de itemi dihotomici sau 50 de itemi cu alegere multiplă ori cu răspuns lacunar. În același interval de

timp se pot rezolva 5-6 itemi cu răspunsuri deschise, care necesită, pentru fiecare răspuns, cam jumătate de pagină (Aiken, 1994).

În orice caz, o testare nu are voie să depășească o oră pentru elevii din școala primară și o oră și jumătate pentru cei din gimnaziu și din liceu. Întrucât copiii preșcolari și elevii din clasele primare nu pot să rămână atenți la sarcinile testului mai mult de 30 de minute, dacă rezolvarea testului necesită un timp mai lung, se va recurge la administrarea testului în mai multe sesiuni de lucru (Aiken, 1994).

Există diverse posibilități pentru construirea testului. De exemplu:

- se extrag aleator itemii din bancă

sau

- dacă a fost întocmită o tabelă de specificații, se respectă frecvențele înscrise în fiecare celulă a sa

sau

- dacă dificultatea itemilor a fost apreciată înainte de introducerea lor în banca de itemi, se aleg itemii astfel încât valorile dificultăților lor să fie repartizate cât mai uniform pe intervalul valorilor posibile sau dorite (uneori se urmărește alcătuirea unui test dificil sau a unui ușor).

Atunci când testul format conține itemi de mai multe tipuri, este recomandabil ca toți itemii de același tip să fie grupați, pentru ca subiectul să nu mai facă și efortul de a identifica, la fiecare item, modul în care trebuie să răspundă. Acest procedeu facilitează, de asemenea, administrarea și cotarea testului.

Aranjarea itemilor în ordinea crescătoare a dificultății îi ajută pe unii subiecți, crescându-le motivația pentru rezolvarea testului: văzând că pot rezolva primii itemi, ei vor anticipa obținerea succesului în continuare și vor avea curajul de a aborda și itemii mai dificili. La testele administrate cu limită de timp și la cele foarte dificile, acest mod de aranjare are ca efect creșterea scorurilor testului. Dar, se pare că efectul său este scăzut sau nul în cazul itemilor cu alegere multiplă.

Se va decide apoi cum vor fi aranjați itemii în pagină, pentru a putea fi ușor de citit, dar și pentru a permite utilizarea economică a hârtiei. În funcție de locul unde vor fi marcate răspunsurile (pe broșura testului sau pe foaia de răspuns) se va stabili așezarea în pagină a pozițiilor care vor conține răspunsurile, astfel încât corectarea să se facă ușor, eventual cu ajutorul unui șablon sau al calculatorului.

După alcătuirea testului, se vor redacta instrucțiunile de administrare. Acestea trebuie să fie clare și să furnizeze suficiente detalii, pentru a nu fi necesare comentarii suplimentare în momentul testării.

Instrucțiunile de lucru vor conține informații privind scopul testării, modul de lucru, din câți itemi se compune testul, cât timp este alocat pentru răs-

puns și cum se cotează răspunsurile. De asemenea, se va menționa dacă este sau nu recomandabil să se ghicească răspunsul, atunci când nu se cunoaște răspunsul corect sau cel mai potrivit, și dacă este sau nu permis să se omită răspunsul.

Autorul testului are sarcina de a utiliza cele mai potrivite elemente grafice (sublinieri, caractere mai mari, culori contrastante) pentru a atrage atenția asupra anumitor părți din instrucțiunile de administrare, așa încât cei care le vor citi subiecților să știe cum să își modifice intonația.

12.5. ADMINISTRAREA TESTULUI

Administrarea testului se va face în aceleași condiții în care va fi utilizat în practică. Se obișnuiește ca broșura testului și foaia de răspuns să le fie înmânate subiecților doar după citirea instrucțiunilor de administrare a testului și rezolvarea eventualelor nelămuriri. Mai întâi se dă foaia de răspuns și se repetă instrucțiunile privind modul de completare a răspunsurilor. Apoi, fiecare subiect va primi broșura testului, așezată cu fața în jos. Ea va fi întoarsă doar la comandă, de toate persoanele în același timp.

Este foarte important să li se spună subiecților ce informații de identificare trebuie să completeze (nume, sex, vârstă, funcție etc.) și unde.

În general (nu numai în faza de construire a testului), dacă testele se administrează în clasă, elevii trebuie să fie informați din timp când și unde se va face testarea, ce fel de test se va administra (oral, scris etc.) și la ce se referă acesta, pentru a se putea pregăti intelectual, emoțional și fizic.

12.6. ANALIZA REZULTATELOR TESTĂRII

Se face o analiză atentă a modului în care a decurs testarea, urmărindu-se dacă instrucțiunile de administrare au fost complete și dacă timpul de lucru a fost suficient. Se prelucrează apoi răspunsurile primite de la subiecți și se încercă să se construiască un test, pe baza celui experimentat. Se procedează astfel:

- (1) Se face analiza de itemi.
- (2) Se elimină itemii necorespunzători.
- (3) Eventual, dintre itemii rămași se selectează cei care satisfac anumite condiții, prin care se asigură o valoare cât mai bună a unui anumit parametru al testului.

(4) Se analizează apoi caracteristicile psihometrice (fidelitatea, validitatea, dificultatea, puterea de discriminare) ale testului obținut.

12.6.1. Analiza de itemi

Analiza de itemi are ca obiective esențiale descifrarea mecanismelor cognitive aplicate de subiecți pentru formularea răspunsurilor la itemi și verificarea calităților itemilor ca instrumente de măsură sau de predicție. Ea furnizează informațiile pe baza cărora se selectează itemii care vor intra în componența testului.

Este important de înțeles rolul suplimentar al analizei de itemi în cazul testelor care măsoară constructe. Dacă la început se dispune de o descriere, mai mult sau mai puțin detaliată a constructului, care servește ca bază pentru scrierea itemilor, după analizarea itemilor se fac inferențe asupra constructului pornind de la relațiile constatate între răspunsurile la itemi. Deci, deși construirea testului începe cu definirea teoretică a constructului, constructul ajunge să fie o inferență fundamentată pe consistența răspunsurilor la itemi (Steinberg, Thissen, 1996).

Alegerea procedurilor statistice aplicate în cadrul analizei de itemi și interpretarea rezultatelor obținute sunt determinate de tipul itemilor, de funcția pe care o va îndeplini testul și de calitățile psihometrice pe care trebuie să le aibă acesta.

Pentru fiecare item, se efectuează una sau mai multe dintre următoarele prelucrări:

- a. calculul frecvenței cu care au fost alese răspunsurile și identificarea cauzelor care au determinat frecvențe foarte mari sau foarte mici;
- b. inspectarea textului itemului pentru depistarea termenilor ambigui, a absenței unor informații necesare pentru formularea unui răspuns corect și identificarea răspunsurilor propuse neplauzibile (pe care nu le alege nimeni) sau evidente (pe care le aleg toți cei testați);
- c. verificarea comportării itemului față de diverse criterii;
- d. determinarea capacității itemului de a face deosebire între examinații care au nivel scăzut și cei care au nivel ridicat la variabila măsurată sau prezisă de test.

Vom descrie în continuare procedeele statistice ale analizei de itemi. Pentru simplificarea expunerii, vom desemna prin expresia "rezolvare corectă/greșită a itemului" indicarea de către subiect a aceluia răspuns care corespunde nivelului înalt/scăzut al variabilei măsurate sau prezise de test.

12.6.1.1. Indicele de dificultate al unui item

Cea mai obișnuită măsură a dificultății itemului este frecvența relativă a persoanelor care au rezolvat corect itemul, din totalul celor cărora li s-a administrat testul (Murphy, Davidshofer, 1987).

Indicele de dificultate astfel definit se notează cu p și reprezintă scorul mediu care s-ar obține la item dacă s-ar acorda un punct pentru răspunsul corect și zero puncte în cazul răspunsului greșit sau omis. Se observă că p este cuprins întotdeauna între 0 și 1.

Davis (1955) recomandă ca, atunci când testul este administrat cu limită de timp și există persoane care nu au reușit să îl parcurgă integral, la calculul frecvenței relative p pentru un anumit item să se utilizeze drept numitor numărul examinațiilor care au citit întrebarea și nu efectivul lotului căruia i s-a administrat testul. Pentru fiecare persoană vor fi considerate necitite toate întrebările aflate după ultimul item rezolvat.

Indicele de dificultate este o caracteristică atât a itemului, cât și a populației căreia i s-a administrat testul. Un item cu răspunsuri la alegere care cere rezolvarea unei ecuații de gradul al doilea va fi rezolvat corect de majoritatea elevilor din clasa a X-a, în timp ce elevii din clasa a III-a vor omite răspunsul sau vor alege unul greșit, deci indicele său de dificultate va fi diferit de la o populație la alta.

Valorile extreme ale lui p restrâng variabilitatea scorurilor la test. Itemii cu $p=0$ (care nu au fost rezolvați corect de nici o persoană) și cei cu $p=1$ (care au fost rezolvați corect de toți cei examinați) nu contribuie la măsurarea variabilei vizate de test și, prin urmare, sunt inutili.

Dispersia scorurilor la un test format din itemi necorelați liniar între ei este maximă atunci când toți itemii au indicele de dificultate egal cu 0.5.

În general, itemii al căror indice de dificultate ia valori în intervalul $[0.3, 0.7]$ permit o diferențiere bună între subiecți (Gregory, 1992).

Pentru itemii dihotomici și pentru cei cu alegere multiplă care au un singur răspuns corect se recomandă să se păstreze în test itemii al căror indice de dificultate este apropiat de valoarea $0.5+g/2$, unde g reprezintă probabilitatea de a ghici răspunsul corect, adică este egal cu raportul 1/ numărul de răspunsuri posibile (Gregory, 1992).

Atunci când testul va fi utilizat pentru selecția persoanelor care vor face parte dintr-un grup extrem (cu valori foarte mari sau foarte mici ale caracteristicii măsurate de test), se va forma testul din itemi care au indicele de dificultate mai mic decât 0.3 sau mai mare decât 0.7.

Pentru întrebările cu alegere multiplă din testele de aptitudini sau de cunoștințe, care pretind selectarea unui singur răspuns, dacă se acceptă

presupunerea că examinații care nu cunosc răspunsul corect aleg la întâmplare dintre cele n răspunsuri propuse, se poate calcula indicele de dificultate cu următoarea formulă:

$$p_c = \frac{np-1}{n-1} \quad (12.6.1.1.1)$$

unde

p este frecvența relativă a persoanelor care au răspuns corect la item.

p_c reprezintă scorul mediu care s-ar obține la item dacă s-ar acorda un punct pentru răspunsul corect și $-1/(n-1)$ puncte pentru răspunsurile greșite sau omise.

Guilford (1965) propune să se utilizeze această formulă atunci când trebuie să se compare, în ceea ce privește dificultatea, itemi care au numere diferite de răspunsuri propuse. Intuitiv, dacă doi itemi se referă la același lucru, sunt bine formulați, așa încât răspunsurile propuse sunt la fel de atractive pentru cei care nu cunosc răspunsul corect, dar unul propune n_1 răspunsuri, iar celălalt n_2 răspunsuri, $n_1 < n_2$, ar fi de așteptat ca primul item să fie rezolvat corect de mai mulți subiecți decât al doilea. Dacă, totuși, frecvența relativă a rezolvărilor corecte este aceeași pentru ambii itemi, înseamnă că primul item este mai dificil. Acest lucru este pus în evidență de indicele p_c , după cum rezultă din inegalitatea:

$$\frac{n_1 p - 1}{n_1 - 1} \leq \frac{n_2 p - 1}{n_2 - 1} \quad (12.6.1.1.2)$$

Indicii de dificultate definiți până acum nu sunt potriviți în cazul întrebărilor care cer mai multe răspunsuri, întrucât nu iau în considerare decât trei situații: item rezolvat integral corect, item greșit și item omis. Rezolvările parțial corecte sunt considerate greșite, indiferent câte răspunsuri corecte s-au dat din cele k cerute. Pentru a evita acest inconvenient și a putea compara asemenea itemi, se utilizează următoarea formulă de calcul al indicelui de dificultate:

$$p_m = \frac{1}{k} \sum_{i=1}^k i p_i \quad (12.6.1.1.3)$$

unde:

p_i = frecvența relativă a celor care au dat i răspunsuri corecte.

p_m reprezintă scorul mediu al itemului dacă s-ar acorda $1/k$ puncte pentru fiecare răspuns corect ales.

Pentru $k=1$, p_m coincide cu p , frecvența relativă a celor care au rezolvat corect itemul.

p_m astfel definit variază între 0 (când nu a fost indicat nici măcar un răspuns corect) și 1 (când toate persoanele au indicat cele k răspunsuri corecte). p_m are valoare mare atunci când frecvența relativă a celor care au dat toate răspunsurile corecte este mare.

În cazul itemilor cu alegere multiplă este bine să se calculeze și frecvența relativă cu care a fost ales fiecare răspuns propus, pentru a identifica răspunsurile corecte care au fost alese de foarte puțini subiecți și răspunsurile greșite alese de multe persoane.

Există situații, cum este cazul scalelor Likert, în care itemul pretinde alegerea unui singur răspuns, dar răspunsurile propuse nu sunt dihotomizate în "corect"/ "greșit" sau "adevărat"/ "fals", ci fiecare răspuns primește un alt punctaj. Atunci, în locul indicelui de dificultate al itemului se vor calcula scorul mediu și abaterea standard a scorurilor itemului. În mod obligatoriu se va analiza și forma repartiției scorurilor itemului, pentru a depista cazurile când sunt alese cu frecvențe mari două sau mai multe răspunsuri care ar trebui să exprime lucruri foarte deosebite între ele și a determina cauza.

O imagine mai clară asupra modului în care persoanele examinate au rezolvat itemul o oferă energia informațională a itemului, calculată cu formula (Cresin, 1966):

$$E = \frac{\sum_{i=1}^k \left(\frac{p_i}{1-o} \right)^2 - \frac{1}{k}}{1 - \frac{1}{k}} \quad (12.6.1.1.4)$$

unde:

p_i = frecvența relativă a celor care au ales răspunsul i , $i=1,2,\dots,k$.

o = frecvența relativă a celor care nu au răspuns la item.

Energia informațională variază între 0 (când toate persoanele au răspuns la item și toate frecvențele relative p_i sunt egale) și 1 (când toate persoanele au ales același răspuns). Ea este mică atunci când toate răspunsurile propuse au fost alese cu frecvențe apropiate și este mare atunci când persoanele au ales un număr mic de răspunsuri.

12.6.1.2. Analiza răspunsurilor incorecte

La o întrebare cu răspunsuri la alegere dintr-un test de aptitudini sau de cunoștințe este important să se verifice dacă factorii de distragere (răspunsurile incorecte din lista de răspunsuri propuse) își îndeplinesc într-adevăr misiunea.

Un procedeu de lucru recomandat în cazul întrebărilor care cer selectarea unui singur răspuns din n răspunsuri propuse este următorul (de Landsheere, 1975; Murphy, Davidshofer, 1987):

a. Pentru fiecare răspuns greșit i , se calculează:

• frecvența observată, f_{oi} , egală cu frecvența relativă a cazurilor în care a fost ales:

$$f_{oi} = \frac{N_i}{N} \quad (12.6.1.2.1)$$

și

• frecvența teoretică, f_t , egală cu frecvența relativă a cazurilor în care ar fi fost ales dacă toți cei care au dat un răspuns greșit ar fi ales răspunsul greșit la întâmplare:

$$f_t = \frac{N - N_c}{N(n-1)} \quad (12.6.1.2.2)$$

unde:

N = numărul de subiecți care au răspuns la item;

N_c = numărul de subiecți care au ales răspunsul corect;

N_i = numărul de subiecți care au ales răspunsul i .

Toate răspunsurile greșite ale itemului au aceeași frecvență teoretică.

b. Se apreciază că întrebarea este bine construită dacă pentru fiecare răspuns greșit frecvența observată f_{oi} este aproximativ egală cu frecvența teoretică f_t .

Exemplu

Tabelul următor prezintă frecvențele de răspuns la un item care propune 4 răspunsuri:

Răspuns	Frecvența alegerii fiecărui răspuns	Frecvența observată a fiecărui răspuns
1	58	0.29
2	4	0.02
corect 3	108	-
4	30	0.15

Frecvența teoretică a fiecărui răspuns greșit este egală cu

$$f_t = \frac{200-108}{200*3} = 0.153$$

Răspunsul greșit 4 are frecvența observată apropiată de frecvența teoretică. În schimb, răspunsurile 1 și 2 au fost alese de mult mai mulți, respectiv mult mai puțini subiecți decât era de așteptat, dacă s-ar fi ghicit răspunsul. Ar trebui analizate răspunsurile greșite pentru a se depista cauzele abaterilor frecvențelor observate de la frecvența teoretică.

Se consideră că un răspuns greșit pe care nu îl aleg decât foarte puține persoane nu își îndeplinește rolul de factor de distragere întrucât este identificat cu ușurință drept greșit și, în acest fel, scade dificultatea sarcinii subiectului, care nu mai are de ales răspunsul corect dintre cele n răspunsuri propuse, ci dintre $n-1$. Un asemenea răspuns trebuie înlocuit.

Un răspuns greșit ales de un număr mare de persoane poate conține o formulare incorectă sau incompletă, care îl face să semene cu răspunsul corect, dar, în același timp, poate semnaliza existența unei erori la răspunsul corect, care îi determină pe subiecți să îl respingă și să aleagă unul dintre celelalte răspunsuri.

Există și situații în care unele răspunsuri greșite sunt formulate intenționat astfel încât să se deosebească de răspunsul corect doar prin câte un element subtil. Ele vor fi alese frecvent, dar întrebările de acest tip sunt utile pentru a-i identifica pe subiecți cu un nivel înalt al variabilei măsurate de test (de exemplu pentru selecția elevilor în lotul olimpic).

La baza algoritmului descris mai sus a stat următorul raționament: Subiecții cărora li se prezintă o întrebare cu alegere multiplă se împart în două categorii; unii cunosc răspunsul corect și îl marchează, ceilalți aleg la întâmplare din lista de răspunsuri propuse. În realitate, însă, modul de selecție a răspunsului diferă de la un subiect la altul. Unele persoane aleg un răspuns eronat cu

mai multe răspunsuri greșite și să le elimine, alegând la întâmplare dintre celelalte. În sfârșit, mai sunt și persoane care aleg răspunsul bazându-se pe alte considerente, cum ar fi lungimea răspunsurilor propuse, precizia limbajului folosit în formularea acestora, poziția răspunsului corect în întrebările pe care le-a parcurs deja etc. Din aceste motive, în multe cazuri răspunsurile greșite sunt alese cu frecvențe diferite unele de altele.

Pornind de la ideile teoriei răspunsului la itemi, Murphy și Davidshofer (1987) recomandă construirea funcției caracteristice a fiecărui răspuns propus. Această funcție are ca valoare, la fiecare nivel al caracteristicii măsurate de test, probabilitatea ca răspunsul respectiv să fie ales. Dacă itemul este bine construit, atunci funcția caracteristică a răspunsului corect va fi nedescrescătoare, în timp ce funcțiile caracteristice ale răspunsurilor greșite vor fi necrescătoare; nici una dintre funcțiile caracteristice nu va fi constantă pe întreaga mulțime de definiție.

Exemplu

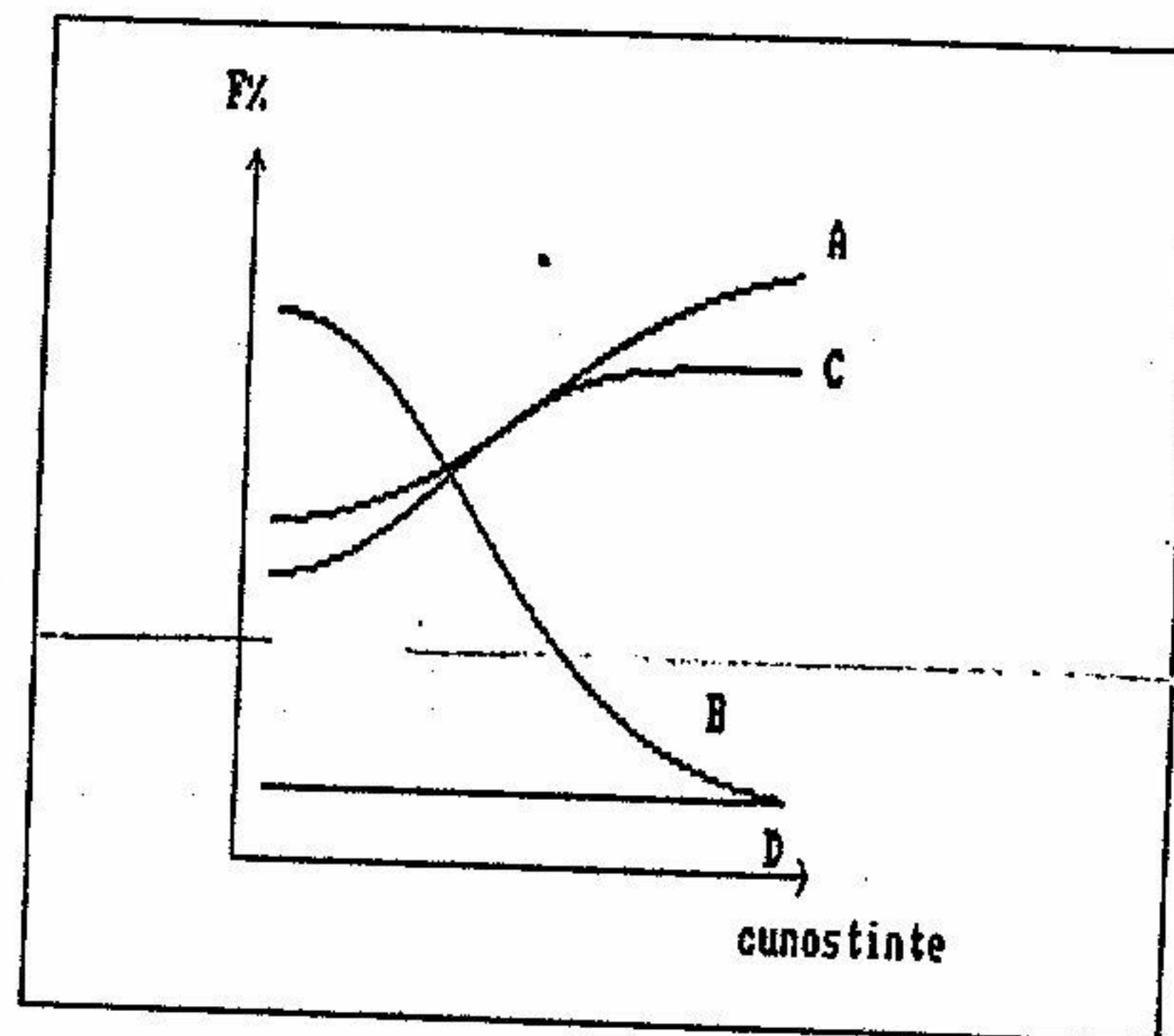


Figura 12.6.1.2.1
Curbele caracteristice ale răspunsurilor la un item cu alegere multiplă

În figura 12.6.1.2.1 sunt prezentate funcțiile caracteristice ale răspunsurilor A (corect), B, C și D ale unui item cu alegere multiplă dintr-un

test de cunoștințe. Se observă că numai răspunsul greșit B are funcția caracteristică descrescătoare. Răspunsul C este ales mai frecvent de subiecți bine pregătiți; este posibil ca el să nu fie formulat clar, deci trebuie verificat. Răspunsul D este ales foarte rar și frecvența alegerii sale nu depinde de nivelul cunoștințelor celor testați. El este inefficient ca factor de distragere și trebuie eliminat.

12.6.1.3. Validitatea itemilor

În general, pentru aprecierea validității itemilor se apelează la un criteriu extern, de exemplu la scorurile unui test care măsoară sau prezice același lucru ca și testul analizat.

Se spune că un item este valid, dacă el poate fi folosit la evaluarea caracteristicii măsurate de variabila criteriu.

De obicei, pentru estimarea validității se calculează coeficientul de corelație între scorurile itemului și valorile variabilei criteriu. Acesta este denumit "indice de validitate" al itemului. Presupunând că variabila criteriu este continuă, tipul coeficientului de corelație utilizat va depinde de tipul itemului, mai precis de numărul scorurilor posibile ale itemului.

În cazul itemilor dihotomici, cărora li se acordă un punct dacă sunt rezolvați corect și zero puncte dacă sunt rezolvați greșit sau nerezolvați, validitatea poate fi măsurată cu ajutorul coeficientului de corelație punct-biserială:

$$r_{bs} = \frac{M_C - M_G}{\sigma_y} \sqrt{pq} \quad (12.6.1.3.1)$$

unde:

p = frecvența relativă a celor care au rezolvat corect itemul, egală cu scorul mediu al itemului;

q = $1-p$

σ_y = abaterea standard a valorilor variabilei criteriu;

M_C, M_G = media valorilor variabilei criteriu pentru cei care au rezolvat itemul corect, respectiv pentru cei care au greșit sau nu l-au rezolvat.

Corelația punct-biserială este un caz particular al corelației produs-moment. Ea corespunde situației când una dintre variabile ia doar două valori.

Dacă testul este administrat cu limită de timp, atunci vor fi luate în considerare pentru calculul corelației numai persoanele care au parcurs itemul.

Verificarea semnificației statistice a coeficientului de corelație punct-biserială se face folosind testul t (Student):

$$t = \frac{r_{bis}}{\sqrt{1-r_{bis}^2}} \sqrt{N-2} \quad (12.6.1.3.2)$$

unde N reprezintă numărul persoanelor care au parcurs itemul.

Itemul este considerat valid, dacă $t > t_{N-2, \alpha}$ fiind valoarea citită din tabelele legii t , care corespunde pragului de semnificație ales α și la $N-2$ grade de libertate.

Se mai poate studia validitatea itemului prin analizarea diferenței mediilor M_C și M_G . Presupunând că în cele două grupe determinate de valorile scorului la item (cei cu scorul 1 și cei cu scorul 0) variabila criteriu este repartizată normal și are aceeași dispersie, se calculează valoarea t folosind formula:

$$t_{df} = \frac{M_C - M_G}{\sigma_y} \sqrt{Npq} \quad (12.6.1.3.3)$$

și se compară cu valoarea $t_{N-2, \alpha}$ citită din tabele. Dacă $t_{df} > t_{N-2, \alpha}$, atunci se consideră că itemul este valid.

Pentru itemii al căror scor poate lua mai mult de două valori, validitatea poate fi estimată cu ajutorul coeficientului de corelație liniară între scorurile la item și valorile variabilei criteriu.

Interpretarea valorii acestui coeficient se face la fel ca în cazul coeficientului de corelație punct-biserială, utilizând testul t .

Atunci când este măsurată printr-un coeficient de corelație liniară, validitatea itemului arată cât de puternică este legătura liniară dintre scorurile itemului și valorile variabilei criteriu.

Uneori este foarte greu sau chiar imposibil de găsit o variabilă criteriu pentru validarea itemilor unui test. În cazul testelor de cunoștințe asemenea situații sunt frecvente.

Dar, dacă specialiștii apreciază că un test, în ansamblu, este valid, adică măsoară cunoștințele din domeniul pentru care a fost construit, atunci scorurile testului pot fi utilizate, în locul valorilor unei variabile criteriu, la validarea itemilor care compun testul. Este adevărat că rezultatele vor fi distorsionate de faptul că în calculul scorului total al testului se ia în considerare și scorul itemului, dar, dacă numărul itemilor din test este mare, influența fiecărui item

este redusă. Vor fi considerați valizi itemii care măsoară același lucru ca și întregul test. În acest caz denumirea de indice de validitate este înlocuită cu aceea de **indice de discriminare și consistență internă** (Davis, 1955). Totuși, *Standards for Educational and Psychological Tests* (1974) consideră că asemenea coeficienți nu măsoară validitatea itemilor.

12.6.1.4. Indicele de discriminare al unui item

Noțiunea de putere de discriminare a unui item vine să o completeze pe cea de validitate a itemului. Pentru estimarea ambelor este necesară existența unei variabile criteriu. Uneori - mai frecvent în cazul testelor de cunoștințe - drept variabilă criteriu poate servi testul din care face parte întrebarea.

În general, se apreciază că un item are putere de discriminare dacă examinații cu valori mari la variabila criteriu au șanse mai mari să rezolve itemul decât cei care au valori scăzute. O asemenea comportare este de așteptat în cazul itemilor valizi, ale căror scoruri sunt corelate liniar și pozitiv cu valorile variabilei criteriu. Din acest motiv, în literatura de specialitate problema validității itemilor și cea a puterii lor de discriminare se confundă uneori, deși metoda utilizată pentru aprecierea puterii de discriminare nu garantează întotdeauna validitatea itemului.

O posibilitate de evaluare a puterii de discriminare a unui item o constituie *metoda grupelor contrastante* (Murphy și Davidshofer, 1987), aplicabilă atunci când testul a fost administrat unui număr mare de persoane. Din mulțimea celor examinați se formează două loturi:

- grupa "bună", compusă din cei care au cele mai mari valori la variabila criteriu;
- grupa "slabă", alcătuită din cei care la variabila criteriu au obținut valorile cele mai mici.

Dacă valorile variabilei după care se face împărțirea în grupele contrastante sunt repartizate conform legii normale, grupele vor conține câte 27% din totalul persoanelor. Dacă repartiția frecvențelor valorilor se abate mult de la legea normală, se vor forma grupele din câte 33% dintre persoane. În general, procentele persoanelor incluse în grupa bună și în grupa slabă vor fi aproximativ egale, fiind cuprinse în intervalul [25, 33] (M.J. Allen și W.M. Yen, după Gregory, 1992).

Pentru fiecare item se calculează indicii de discriminare:

$$D = \frac{C_B}{N_B} - \frac{C_S}{N_S} \quad (12.6.1.4.1)$$

unde:

N_B, N_S = numărul de persoane în grupa bună, respectiv în grupa slabă. Dacă testul a fost administrat cu limită de timp, atunci N_B și N_S vor reprezenta numărul de persoane care au parcurs itemul, în fiecare din cele două grupe;

C_B, C_S = numărul persoanelor care au rezolvat itemul corect, în grupa bună, respectiv în grupa slabă.

Deci, indicele de discriminare este diferența a două frecvențe relative. El variază între -1 (când toți cei din grupa slabă au rezolvat itemul corect și toți cei din grupa bună au greșit) și +1 (când itemul a fost rezolvat corect de toți cei din grupa bună și greșit de toți cei din grupa slabă). Cu cât indicele de discriminare este mai ridicat, cu atât itemul reușește mai bine să diferențieze între cei "buni" și cei "slabi".

Valoarea indicelui de discriminare, D , este dependentă de cea a indicelui de dificultate, p . Atunci când toți examinații rezolvă corect itemul ($p=1$) sau când nimeni nu răspunde corect ($p=0$), itemul nu diferențiază între cei buni și cei slabi. Dacă valoarea p este apropiată de 0 sau de 1 capacitatea de discriminare a itemului este redusă. Cel mai mare potențial pentru a fi buni discriminatori îl au itemii al căror indice de dificultate este apropiat de 0.5.

Ori de câte ori se obține o valoare negativă a indicelui de discriminare la un item cu alegere multiplă, este necesar să se revadă răspunsurile propuse. Când unul sau mai multe dintre răspunsurile greșite par foarte plauzibile unui examinat care are un nivel înalt al variabilei măsurate de test (inteligență, cunoștințe etc.) și când recunoașterea răspunsului corect depinde de un element extrem de subtil, este posibil ca examinații cu un nivel mediu al variabilei evaluate să fie penalizați. La un test de cunoștințe, de exemplu, cei care nu cunosc nimic și aleg la întâmplare, având șanse să nimerească răspunsul corect, pot să obțină un punctaj mai bun decât cei care au cunoștințe parțiale și sunt derutați de modul în care sunt formulate răspunsurile propuse.

În cazul întrebărilor dihotomice, atunci când numărul persoanelor examinate este mai mare de 20 și efectivul celor două grupe este același ($N_B=N_S$), valoarea indicelui D permite următoarele aprecieri (Guilbert, 1987):

$D \geq 0.35$	item excelent;
$0.25 \leq D < 0.35$	item bun;
$0.15 \leq D < 0.25$	item la limită - de revizuit;
$D < 0.15$	item slab - de reexaminat atent și, probabil, de eliminat.

Un item poate avea un indice de discriminare convenabil, fără a avea o validitate bună. Un asemenea caz corespunde următoarei situații: un item dintr-un test de cunoștințe pare ușor celor care au un nivel mediu al cunoștințelor; în realitate, el conține o greșeală de formulare care face ca două dintre răspunsurile propuse să fie la fel de plauzibile. Examinații bine pregătiți au dubii în legătură cu răspunsul corect și din acest motiv greșesc mai frecvent decât ceilalți. Curba de regresie a scorurilor la item față de scorurile variabilei criteriu este prezentată în figura 12.6.1.4.1.

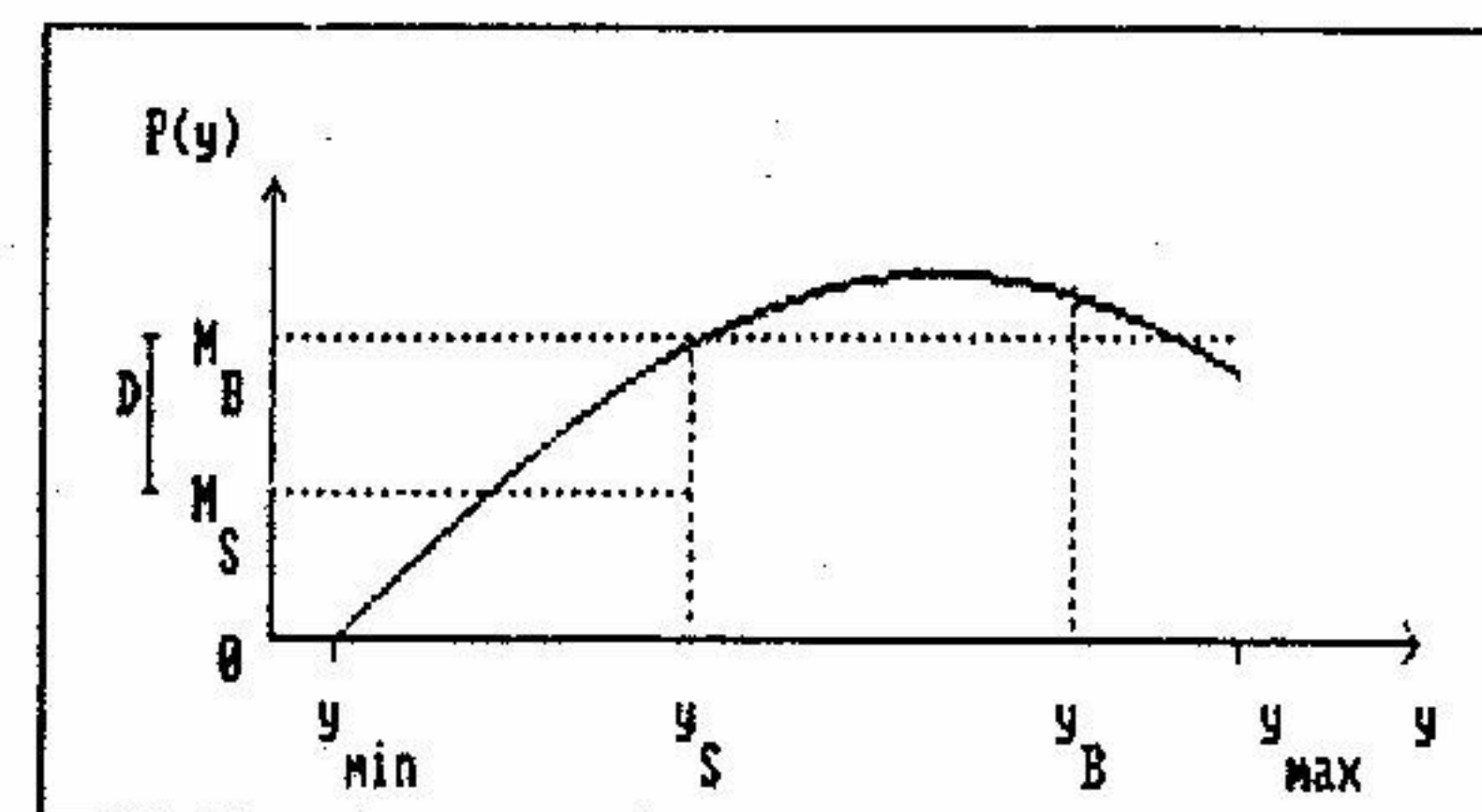


Figura 12.6.1.4.1

Un item care discriminează bine dar nu este valid

Ținând cont de faptul că indicele de discriminare se exprimă cu ajutorul a două frecvențe relative, pentru interpretarea corectă a valorii sale se poate aplica un test statistic. Astfel, dacă efectivele celor două grupe sunt mari ($N_B > 30$ și $N_S > 30$), atunci compararea frecvențelor relative $p_B = C_B/N_B$ și $p_S = C_S/N_S$ se poate face cu ajutorul testului z . Se calculează valoarea z cu formula:

$$z = \frac{p_B - p_S}{\sqrt{p(1-p)\left(\frac{1}{N_B} + \frac{1}{N_S}\right)}} \quad (12.6.1.4.2)$$

unde:

$$p = \frac{C_B + C_S}{N_B + N_S} \quad (12.6.1.4.3)$$

și se compară cu valoarea z_α citită din tabelul legii normale, corespunzătoare pragului de semnificație α ales. Dacă $z > z_\alpha$ atunci se apreciază că puterea de discriminare a itemului este satisfăcătoare.

O deficiență a metodei prezentate pentru calculul indicelui de discriminare constă în faptul că elimină din prelucrare mai mult de o treime din totalul celor examinați. În plus, indicele de discriminare astfel calculat nu ne spune cum diferențiază itemul între persoanele din fiecare grupă, ci doar dacă, în cadrul întregului eșantion, face deosebire între cei mai buni și cei mai slabi.

Analog indicelui de discriminare D , utilizat pentru itemii dihotomici, se poate defini un indice de discriminare valabil în cazul itemilor al căror scor admite mai multe valori:

$$D_m = M_B - M_S \quad (12.6.1.4.4)$$

unde M_B și M_S reprezintă media scorurilor itemului în grupa bună, respectiv în cea slabă.

Această formulă este mai generală decât cea a indicelui D ; dar, dacă itemul este dihotomic, atunci D_m coincide cu D .

Se consideră că itemul este satisfăcător în privința puterii sale de discriminare dacă diferența mediilor M_B și M_S este semnificativă statistic. Pentru compararea mediilor se poate utiliza testul t . Presupunând că scorurile itemului sunt repartizate normal și au aceeași dispersie în cele două grupe, se calculează:

$$t = \frac{M_B - M_S}{\sqrt{\frac{(N_B - 1)\sigma_B^2 + (N_S - 1)\sigma_S^2}{N_B + N_S - 2}}} \sqrt{\frac{N_B + N_S}{N_B N_S}} \quad (12.6.1.4.5)$$

unde σ_B^2 și σ_S^2 reprezintă dispersia scorurilor itemului în grupa bună, respectiv în grupa slabă.

Dacă t este mai mare decât valoarea $t_{m, \alpha}$ citită din tabelele legii t pentru $m = N_B + N_S - 2$ grade de libertate și pragul de semnificație α , se consideră că itemul are o putere de discriminare satisfăcătoare.

Indicele de discriminare D_m , la fel ca și D , arată doar în ce măsură o întrebare diferențiază între o grupă bună și una slabă. El nu spune nimic despre felul în care itemul discriminează între indivizi care au niveluri diferite ale variabilei măsurate sau prezise de test.

12.6.1.5. Indicele de fidelitate al itemului

Acest indice servește la aprecierea contribuției pe care și-o aduce itemul la asigurarea unei bune consistențe interne a testului.

Conform formulei (8.4.2.2), indicele de fidelitate este dat de produsul dintre abaterea standard a scorurilor sale, σ_y , și coeficientul de corelație liniară dintre scorurile itemului și scorurile testului, $\rho(Y, X)$:

$$Rel(Y) = \sigma_y \rho(Y, X)$$

Din formula de calcul a coeficientului α al lui Cronbach, (8.4.2.3):

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum_{i=1}^n \sigma_i^2}{\sigma_X^2 \sum_{i=1}^n Rel(Y_i)} \right)$$

se observă că testele alcătuite din itemi cu valori mari ale indicilor de fidelitate au o consistență mare.

Este dificil de apreciat contribuția pe care și-o aduce dispersia scorurilor itemului la mărimea coeficientului de consistență internă, întrucât în expresia coeficientului α abaterea standard a scorurilor itemului, σ_y , apare atât la numărătorul, cât și la numitorul raportului care se scade din 1. Dar, pentru obținerea unei valori cât mai mari a coeficientului α este important ca itemii care vor fi introduși în test să aibă corelații liniare mari, pozitive, cu scorul total al testului.

12.6.1.6. Observații asupra analizei de itemi

Analiza itemilor unui test se poate face conform următorului algoritm:

(1) Se calculează indicii de dificultate pentru toți itemii și se elimină itemii rezolvați de toate persoanele și cei nerezolvați de nici un subiect.

(2) Se încearcă depistarea cauzelor pentru care unii indici de dificultate sunt foarte mari sau foarte mici și, pe această bază, se elimină itemii cu greșeli.

(3) Pentru itemii cu răspunsuri la alegere, se analizează răspunsurile incorecte și se elimină itemii în care unele răspunsuri greșite propuse au fost alese de foarte mulți sau de foarte puțini subiecți.

(4) Se aplică un procedeu al analizei de itemi determinat de condițiile pe care trebuie să le îndeplinească testul care se construiește:

- Dacă testul trebuie să discrimineze între două grupe contrastante, mai precis, să furnizeze scoruri care să difere mult de la o grupă la alta, atunci se calculează pentru fiecare item indicele de discriminare și se elimină itemii necorespunzători din acest punct de vedere.
- Dacă testul trebuie să se comporte într-un anumit fel față de anumit criteriu (să coreleze liniar pozitiv sau negativ ori, din contră, să nu coreleze semnificativ), atunci se vor elimina itemii care au o altfel de comportare.

(5) Se elimină itemii cu indicele de dificultate necorespunzător scopului pe care trebuie să îl îndeplinească testul: în general, se elimină itemii foarte ușori și cei foarte grei, dar, dacă testul trebuie să identifice persoanele cu niveluri foarte scăzute sau foarte ridicate ale unei caracteristici, se vor reține doar itemii ușori, respectiv dificili.

(6) Se calculează coeficientul de corelație între scorurile testului și scorurile itemului și se elimină itemii cu corelații nesemnificative sau negative.

Din itemii rămași vor fi selectați cei care vor alcătui testul. Pentru testul rezultat se vor studia atât caracteristicile psihometrice (fidelitate, validitate, dificultate, putere de discriminare), cât și lungimea testului, tipurile itemilor componenți și, în funcție de acestea, timpul de rezolvare a testului și costurile administrării sale.

12.6.2. Metode de selectare a itemilor pentru construirea unui test psihologic

Metoda intuitivă (rațională) se utilizează pentru a obține un test care să măsoare cu precizie diferențele individuale în privința unei trăsături comune sau a unui atribut care prezintă interes științific și care nu este măsurat adecvat prin nici o tehnică existentă (Edwards, 1970). Se deduce că această metodă urmărește formarea unui test pentru care sunt importante validitatea relativă la construcție și consistența internă.

Se pornește de la o mulțime de itemi care descriu comportamente specifice constructului la care se referă testul, formulați pe baza intuiției autorului testului. De aici provine denumirea de "metodă intuitivă". În unele lucrări, denumirea de "metodă rațională" se folosește ca un sinonim pentru

"metoda intuitivă", în alte lucrări este considerată "metodă rațională" doar cea care utilizează procedee statistice pentru selectarea itemilor din colecția existentă (Henry, Sanford, 1972)

Pentru a putea aplica formule statistice, itemii sunt administrați unui lot format din 200-400 de persoane (Edwards, 1970) și se calculează coeficientul de consistență internă. Dacă acesta are o valoare prea scăzută se încearcă să se formeze testul dintr-un număr mai restrâns de itemi.

În general se utilizează două căi de selecție a itemilor, ambele bazate pe scorurile acestora în cadrul eșantionului:

- metoda consistenței interne;
- metoda teoretică.

A. Metoda consistenței interne are mai multe variante. Vom prezenta câteva dintre acestea:

- Dacă numărul de itemi este mai mare decât 60, se elimină itemii care au covarianțe scăzute cu scorul total al testului, iar dacă numărul de itemi este mai mic decât 20, se adaugă itemi noi (Edwards, 1970).

- Sunt eliminați itemii care au coeficienți de corelație scăzuți cu scorul total al testului (Henry, Sanford, 1972).

- Sunt îndepărtați itemii care au o putere de discriminare scăzută, grupele contrastante utilizate pentru calculul indicelui de discriminare fiind formate pe baza scorurilor testului (Anastasi, 1976).

Deficiența acestor metode constă în faptul că nu asigură selectarea acelor itemi care ar furniza cea mai mare valoare posibilă a coeficientului de fidelitate. În metodele consistenței interne fiecare item este comparat cu întregul test și este posibil ca un item necorespunzător să provoace eliminarea unor itemi buni.

B. Metoda teoretică (Shackleton, Fletcher, 1984) grupează itemii și reține itemii unei clase omogene.

Pentru grupare se folosește un algoritm al analizei de clusteri.

Prin acest algoritm la fiecare pas se construiește pe mulțimea de itemi o partiție. În cadrul fiecărei partiții, clasa care se referă la constructul măsurat de test trebuie identificată prin analiza conținutului itemilor. Se formează testul din itemii clasei respective și se calculează coeficientul de fidelitate al testului. Deci, la fiecare pas al algoritmului există câte un test. Dintre acestea se alege cel care are o fidelitate bună și este compus dintr-un număr suficient de mare de itemi ca să poată diferenția între subiecți.

În privința modului de validare a testului construit prin metoda rațională, părerile sunt contradictorii.

Edwards (1970) atrage atenția că un asemenea test nu trebuie corelat cu nici un criteriu, deoarece un test se construiește prin metoda rațională doar atunci când nu există nici o altă posibilitate de a măsura trăsătura la care se referă acesta.

Anastasi (1976) propune să se studieze corelația dintre testul construit și alte teste similare existente, pentru a vedea dacă toate măsoară același lucru. Valorile coeficienților de corelație trebuie să fie mari, dar nu foarte mari, căci dacă se identifică o corelație puternică, fără ca noul test să aibă avantajul de a fi mai ușor de administrat, se poate considera că el nu este decât o dublură a testului existent și nu prezintă nici un interes.

D.T. Campbell (Anastasi, 1976) apreciază că pentru a verifica validitatea relativă la construct a testului trebuie să se studieze corelațiile acestuia cu teste care se referă la variabile psihice despre care se consideră că au legătură cu constructul măsurat de test (validitatea convergentă) și cu teste care vizează variabile psihice despre care se presupune că nu au legătură cu constructul măsurat de test (validitatea discriminantă).

Metoda analizei criteriului intern (metoda analizei factoriale) (Mischel, 1968; Henry, Sanford, 1972; Kline, 1993) servește la construirea unui test destinat să măsoare un anumit factor, care se consideră că reprezintă o variabilă de personalitate fundamentală.

Sunt scriși mai mulți itemi. Aceștia sunt administrați unui lot de subiecți împreună cu alte măsuri (teste, itemi etc.) care au fost identificate în cadrul unor cercetări anterioare ca având legătură cu variabila respectivă. Matricea intercorelațiilor scorurilor obținute pentru toate măsurătorile efectuate este prelucrată prin analiză factorială. Se identifică factorul pe care trebuie să îl măsoare testul și se rețin numai itemii care au saturații mari în factorul respectiv. Dacă este necesar, se schimbă modul de cotare a itemilor astfel încât toate saturațiile să fie pozitive.

Trebuie menționat că factorul comun unei grupări de itemi nu corespunde, în mod obligatoriu, unei caracteristici "reale" sau "primare" a persoanelor examinate (Mischel, 1968). Factorii identificați prin analiză factorială sunt dependenți de itemii prelucrați statistic, de subiecții care au răspuns la test și de metodele aplicate în cadrul analizei factoriale. Variabilele puse în evidență de analiza factorială reflectă doar corelațiile dintre mai multe instrumente de măsură particulare (itemii testului).

Prin metoda analizei factoriale este mai ușor să se construiască deodată mai multe teste decât unul singur. Aceasta deoarece pasul inițial al analizei factoriale are tendința de a evidenția un factor general și mai mulți factori bipolari, iar prin rotația factorilor se reduce varianța factorului general (Kline,

1993). Prin urmare, de la început se vor selecta itemi care să măsoare mai mulți factori. Aceștia vor fi administrați unui lot format cam din 200 de persoane sau, în orice caz, care nu este mai mic de 100 de indivizi. Kline (1993) recomandă ca numărul itemilor prelucrați să fie de două ori mai mare decât cel al itemilor care vor fi păstrați în teste, iar numărul subiecților să fie de 3 ori mai mare decât numărul itemilor introduși în analiza factorială. Astfel, de exemplu, dacă se vor construi 5 scale a câte 20 de itemi, atunci vor trebui administrați $2 \times 5 \times 20 = 200$ de itemi, unui lot format din $3 \times 200 = 600$ de persoane.

În ceea ce privește tipul de rotație utilizat în analiza factorială, se va aplica o rotație ortogonală dacă există argumente teoretice în favoarea ideii că variabilele măsurate de testele care vor fi construite sunt necorelate liniar între ele; în caz contrar, se va aplica o rotație oblică. Pentru fiecare test vor fi selectați numai itemii care au saturații mari într-un singur factor.

Prin această metodă se obțin teste care măsoară câte un singur factor. Pentru a afla ce reprezintă acest factor, se va studia modul în care corelează scorurile testului cu diverse teste despre care se știe ce măsoară.

Această metodă se folosește atunci când se dorește să se obțină teste cu o validitate relativă la construct bună.

Metoda exploratorie (Ozer, Reise, 1994) se folosește pentru construirea testelor destinate să evalueze constructe care la început sunt incomplet înțelese. În paralel cu construirea testului se dezvoltă și teoria care stă la baza sa. De fapt, această metodă transpune în practică spirala inductiv-ipotetico-deductivă a lui Cattell.

Se începe cu o idee nefinisată despre un construct de personalitate și se scriu cât mai mulți itemi referitori la el. Aceștia sunt administrați unui lot de subiecți. Răspunsurile obținute nu sunt folosite pentru a studia și îmbunătăți calitățile psihometrice ale testului, ci pentru a genera teorii noi. Sunt scriși alți itemi, potriviți noii idei despre construct, iar răspunsurile obținute servesc la o nouă evaluare a teoriei. Procesul se repetă până când se ajunge la un nivel satisfăcător de convergență a răspunsurilor. Itemii respectivi vor constitui testul.

Metoda criteriului extern se utilizează de obicei în cazul testelor care trebuie să aibă o validitate predictivă ridicată.

Este foarte important ca itemii introduși în prelucrare să se refere la aspecte cât mai diverse, să formeze o colecție cât mai eterogenă. Alegerea itemilor care vor intra în componența testului se bazează pe relațiile, determinate empiric, dintre itemii testului și o măsură particulară folosită drept criteriu. Din acest motiv, metoda este denumită și "empirică" (Edwards, 1970; Henry, Sanford, 1972).

De obicei se studiază validitatea fiecărui item relativ la acel criteriu și se elimină itemii care nu sunt valizi.

Uneori, criteriul împarte indivizii în două clase contrastante: unii, pentru care o anumită trăsătură, caracteristică sau aptitudine este foarte accentuată, formează grupul "criteriu", iar ceilalți constituie grupul "de control". În acest caz se studiază puterea de discriminare a itemilor și se elimină itemii care nu discriminează între grupurile contrastante.

Cele două grupuri contrastante trebuie să conțină, împreună, cel puțin 200 de persoane, dacă toate sunt de același sex, sau cel puțin 200 de bărbați și 200 de femei. În cel de-al doilea caz, se selectează doar itemii care discriminează între grupurile contrastante pentru ambele sexe.

Adesea, criteriul care stă la baza împărțirii persoanelor în cele două grupuri este multidimensional. Din acest motiv, este posibil ca testul format din itemii reținuți să nu diferențieze suficient de bine între grupuri. Sau este posibil ca fiecare item să se refere la un alt aspect al criteriului; aceasta face ca intercorelațiile itemilor să fie reduse și, deci, consistența internă a testului să fie scăzută. Pentru a înțelege ce măsoară testul (care sunt factorii care discriminează între grupurile contrastante) se pot prelucra prin analiză factorială itemii reținuți împreună cu itemi ai altor teste, care se știe ce măsoară, și apoi se vor identifica factorii obținuți.

Metoda analizei constructului (Oosterveld, Vorst, 1995) se bazează pe corelațiile itemilor cu diverse variabile care, conform teoriei ce stă la baza elaborării testului, au legături sau, din contră, nu trebuie să aibă legături cu acest construct. Itemii ale căror corelații sunt necorespunzătoare se elimină.

Această metodă se aseamănă cu metoda empirică, doar că aici se folosesc mai multe criterii, inclusiv unele cu care itemii trebuie să nu coreleze semnificativ. Aplicarea metodei constructului necesită existența unei rețele nomologice care să includă constructul măsurat de test.

Alte metode de construire a testelor

Algoritmii pe care îi prezentăm în continuare construiesc un test pornind de la o colecție de itemi dihotomici și de la un item fixat, apreciat ca fiind reprezentativ pentru ceea ce testul trebuie să măsoare sau să prezică.

La fiecare pas se adaugă un item testului existent în acel moment. Itemului adăugat (care se alege astfel încât să maximizeze coeficientul de consistență internă, respectiv un coeficient de discriminare între grupele contrastante) i se impune să aibă funcția de regresie față de scorurile testului existent nedescrescătoare. În plus, se ține cont și de următoarea observație:

Este posibil ca fiecare dintre itemii i_1, i_2, \dots, i_n să aibă funcția de regresie monotonă față de testul format din ei, itemul i_{n+1} să aibă funcția de regresie monotonă față de testul alcătuit din itemii $i_1, i_2, \dots, i_n, i_{n+1}$, dar să existe cel puțin un item $i_j, j \in \{1, 2, \dots, n\}$ a cărui funcție de regresie față de testul format din itemii i_1, i_2, \dots, i_{n+1} să nu fie monotonă.

Din acest motiv nu se adaugă un item testului decât după ce se verifică dacă toți itemii existenți în test au funcții de regresie monotone față de testul care s-ar forma prin adăugarea itemului.

Algoritmii aplică următoarele formule de calcul:

- Scorul testului se obține prin însumarea cotelor itemilor componenți.

- Pentru determinarea funcției de regresie a unui item față de test se aplică metoda celor mai mici pătrate folosind polinoame de gradul al III-lea. Concret, dacă testul este alcătuit din n itemi: i_1, i_2, \dots, i_n , scorurile posibile fiind $0, 1, \dots, n$ și dacă pentru fiecare item $i_k (k=1, 2, \dots, n)$ notăm cu p_k funcția:

$$p_k: \{0, 1, \dots, n\} \rightarrow [0, 1]$$

$\left\{ \begin{array}{l} \text{frecvența relativă a persoanelor care obțin un} \\ \text{punct la itemul } i_k \text{ din totalul celor care au} \\ \text{realizat } t \text{ puncte la test, } t=1, 2, \dots, n \end{array} \right.$

$$p_k(t) = \begin{cases} 0, & t=0 \end{cases}$$

atunci se determină funcția de regresie a itemului i_k față de test, c_k , care îndeplinește următoarele condiții:

$$c_k: [0, n] \rightarrow [0, 1]$$

$$c_k \in \{a_0 + a_1x + a_2x^2 + a_3x^3 \mid a_0, a_1, a_2, a_3 \in \mathbb{R}\}$$

$$\sum_{i=0}^n [p_k(i) - c_k(i)]^2$$

ia valoarea minimă.

Se observă că prin condițiile impuse nu s-au eliminat dreptele de regresie. Ele corespund cazului când c_k este un polinom de gradul întâi pe intervalul $[0, n]$. Dar, pentru unii itemi i_k , se poate găsi o funcție ξ care să aproximeze punctele $p_k(0), p_k(1), \dots, p_k(n)$ mai bine decât dreapta de regresie și, în plus, să imite mai bine alura funcției p_k , punând în evidență eventualele schimbări de monotonie, ceea ce polinoamele de gradul întâi nu sunt capabile să facă.

A. Un algoritm de construire a unui test cu o consistență internă ridicată (Albu, Pitariu, 1992b)

Algoritmul pe care îl prezentăm în continuare construiește un test psihologic prin extragerea itemilor dintr-o mulțime de itemi dihotomici M , astfel încât:

- funcția de regresie a fiecărui item al testului este nedescrescătoare față de test;

- la testul existent la un moment dat se adaugă acel item care produce cel mai mare coeficient de fidelitate Kuder-Richardson, ρ_{20} .

Algoritmul se compune din următorii pași:

Pasul 1 : $M := \{i_1, i_2, \dots, i_n\}$ este mulțimea de itemi;

i_0 este itemul de pornire;

$s := \{i_0\}$ este testul construit

$k := 1$ este numărul de itemi ai testului s

Pasul 2 : $P = \emptyset$

Pentru fiecare item $i_j \in M$ se determină funcția de regresie a itemului i_j față de testul $s \cup \{i_j\}$ și, dacă aceasta este nedescrescătoare, se pune $P := P \cup \{i_j\}$.

Pasul 3 : Dacă $P = \emptyset$, se trece la pasul 8.

Pasul 4 : Se pune $KRMAX := 0$ și $JMAX := 0$

Pasul 5 : Pentru fiecare $i_j \in P$:

- se calculează, folosind scorurile testului $s_j := s \cup \{i_j\}$ coeficientul de fidelitate ρ_{20} al testului s_j ; $KR(s_j)$;

- dacă $KR(s_j) > KRMAX$, se pune $KRMAX := KR(s_j)$ și $JMAX := j$

Pasul 6 : Pentru fiecare item $i_{JMAX} \in s \cup \{i_{JMAX}\}$:

- se determină funcția de regresie față de testul $s \cup \{i_{JMAX}\}$;

- dacă aceasta nu este nedescrescătoare, se elimină itemul i_{JMAX} din mulțimea P :

$P := P \setminus \{i_{JMAX}\}$

și se trece la pasul 3.

Pasul 7 : Se adaugă itemul i_{JMAX} testului s și se elimină din M :

$s := s \cup \{i_{JMAX}\}$

$M := M \setminus \{i_{JMAX}\}$

$k := k + 1$

Dacă $M \neq \emptyset$ se trece la pasul 2.

Pasul 8 : STOP

S-a obținut testul s cu k itemi.

Algoritmul poate fi modificat, introducând condiția de oprire dacă valoarea coeficientului de fidelitate se află sub un anumit prag sau atunci când valorile coeficienților ρ_{20} scad de la un pas la altul.

B. Un algoritm de construire a unui test care să discrimineze bine între două grupe contrastante (Albu, Pitariu, 1991b)

Acest algoritm servește la construirea unui test care să discrimineze cât mai bine între două grupe contrastante, G_1 și G_2 , apreciate a fi foarte deosebite în privința variabilei la care se referă testul.

El urmărește ca:

- itemii introduși în test să aibă funcția de regresie față de test nedescrescătoare;

- scorurile medii ale testului în cele două grupe, G_1 și G_2 , să fie cât mai diferite posibil;

- în cadrul fiecărei grupe, scorurile testului să fie cât mai apropiate de scorul mediu al grupeii.

Următoarele două condiții de mai sus au fost formulate matematic cu ajutorul funcției C :

$C : S \rightarrow R$

$$C(s) = \frac{m_1 - m_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (12.6.2.1)$$

unde :

S = mulțimea tuturor testelor care se pot forma cu itemii mulțimii M ;

s = un test din S ;

pentru $i=1,2$:

n_i = volumul grupeii G_i ;

m_i = scorul mediu al testului s în grupa G_i ;

σ_i^2 = dispersia scorurilor testului s în grupa G_i .

Înainte de aplicarea algoritmului se verifică modul de cotare a itemilor, astfel încât fiecare item să aibă media scorurilor mai mare în grupa G_1 .

Algoritmul se compune din următorii pași:

Pasul 1 : $M := \{i_1, i_2, \dots, i_n\}$ este mulțimea de itemi;

$s := \emptyset$ este testul construit

$k := 0$ este numărul de itemi ai testului s

Pasul 2 : Se pune $P := \emptyset$

Pentru fiecare item $i_j \in M$:

- dacă $s = \emptyset$ se pune $P := P \cup \{i_j\}$;

- dacă $s \neq \emptyset$, se determină funcția de regresie a itemului i_j față de testul $s \cup \{i_j\}$ și, dacă aceasta este nedescrescătoare, se pune $P := P \cup \{i_j\}$.

Pasul 3 : Dacă $P=\emptyset$, se trece la pasul 7.

Pasul 4 : Se pune $C_{MAX} := 0$ și $J_{MAX} := 0$

Pentru fiecare item $i_j \in P$:

- se calculează, folosind scorurile testului $s_j := sU\{i_j\}$:

$$C(s_j) = \frac{m_1 - m_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- dacă $C(s_j) > C_{MAX}$, se pune $C_{MAX} := C(s_j)$ și $J_{MAX} := j$

Pasul 5 : Pentru fiecare item $i_{MAX} \in sU\{i_{MAX}\}$:

- se determină funcția de regresie față de testul $sU\{i_{MAX}\}$;
- dacă aceasta nu este nedescrescătoare, se elimină itemul i_{MAX} din mulțimea P :

$$P := P \setminus \{i_{MAX}\}$$

și se trece la pasul 3.

Pasul 6 : Se adaugă itemul i_{MAX} testului s și se elimină din M :

$$s := sU\{i_{MAX}\}$$

$$M := M \setminus \{i_{MAX}\}$$

$$k := k+1$$

Dacă $M \neq \emptyset$ se trece la pasul 2.

Pasul 7 : STOP

S-a obținut testul s cu k itemi.

12.7. EXPERIMENTAREA SUPLIMENTARĂ A TESTULUI

Efectuarea unei verificări suplimentare a testului, folosind un eșantion de persoane asemănător celui folosit pentru construirea sa, este necesară ori de câte ori alcătuirea testului s-a realizat prin selecția itemilor dintr-o bancă de itemi pe baza răspunsurilor date de subiecți (deci itemii nu au fost aleși de experți). Adică, s-a format un test care este "cel mai bun posibil" sau unul dintre "cele mai bune posibil" pentru eșantionul de persoane folosit, ținând cont de itemii care au fost administrați și de criteriul de optim utilizat (de exemplu, s-a urmărit maximizarea fidelității sau a validității relative la criteriu). La formarea testului și-au adus contribuția și diverși factori legați de șansa care a acționat în eșantionul de persoane care a participat la construirea testului dar care s-ar putea

să nu mai fie prezenți în alte ocazii de testare, cu alți subiecți. Din acest motiv trebuie verificată comportarea testului și într-un alt lot de persoane.

În literatura de specialitate se recomandă ca după alcătuirea testului să se efectueze o validare suplimentară (cross-validation). Prin aceasta se verifică, folosind ecuația de regresie a scorurilor testului față de un criteriu, dacă testul prezice criteriul într-un eșantion nou de subiecți la fel de bine ca în eșantionul original (Gregory, 1992). În practică s-a constatat o scădere a valorii coeficientului de validitate prin schimbarea lotului de subiecți (validity-shrinkage).

Alături de studiul validității testului, este bine că în noul eșantion să se cerceteze și fidelitatea, dificultatea și sensibilitatea testului.

În această fază de experimentare suplimentară a testului se verifică nu numai calitățile psihometrice ale acestuia ci și felul în care au fost redactate instrucțiunile de administrare și au fost editați itemii.

Pentru a se observa eventualele deficiențe, este util să li se dea subiecților un chestionar de opinie, pe care să îl completeze imediat după ce au terminat de răspuns la test. Un asemenea chestionar este anonim și conține întrebări la care se răspunde prin "da" sau "nu" și, eventual, se cere justificarea unuia dintre răspunsuri. Exemple de asemenea întrebări sunt următoarele:

- Condițiile de testare (temperatură, lumină, ventilație etc.) au fost satisfăcătoare? (DA/NU)

Dacă răspunsul este NU, explicați ce v-a nemulțumit.

- Instrucțiunile de lucru au fost complete? (DA/NU)

Dacă răspunsul este NU, ce ar fi trebuit spus în plus?

- Instrucțiunile de lucru au fost clare? (DA/NU)

Dacă răspunsul este NU, ce a fost neclar?

- Textul itemilor a fost ușor de înțeles? (DA/NU)

Dacă răspunsul este NU, unde ați avut dificultăți?

- Modul în care a trebuit să completați răspunsurile a fost simplu? (DA/NU)

Dacă răspunsul este NU, ce a fost dificil?

- Timpul de lucru a fost suficient? (DA/NU)

- Ați răspuns la toți itemii? (DA/NU)

Dacă răspunsul este NU, de ce?

- Vi s-a părut bun testul? (DA/NU)

Dacă răspunsul este NU, ce nu v-a plăcut?

12.8. TRADUCEREA ȘI ADAPTAREA UNUI TEST

Pentru aplicarea unui instrument de măsură într-un grup cultural nou nu este totdeauna suficientă traducerea textului itemilor și a instrucțiunilor de administrare. Uneori este necesară adaptarea testului, prin înlocuirea unor itemi care nu se potrivesc noii populații, alteleori, testul trebuie reconstruit. Aceasta, deoarece între populația căreia i-a fost destinat testul inițial și noua populație pot exista deosebiri în ceea ce privește (Van de Vijver, Hambleton, 1996):

a. comportamentele asociate constructului măsurat de test;

De exemplu, pentru constructul "pietate filială", comportamentele asociate cu ceea ce se înțelege prin "a fi un fiu bun" (a avea grijă de părinți, a le satisface cerințele, a-i trata bine) sunt mult mai numeroase în China decât în Europa de Vest.

Se impune cercetarea modului în care este înțeles constructul în viața zilnică și a comportamentelor asociate lui. Se pot identifica astfel itemii nepotriviți noii populații.

b. obișnuința persoanelor cu situația de testare ori cu forma particulară a testului;

Lipsa obișnuinței de a răspunde la teste (în general sau la teste de acest tip) influențează asupra răspunsurilor la un număr mare de itemi sau chiar la toți itemii, producând diferențe ale scorului total al testului între cele două populații.

c. răspunsurile date la unii itemi din test, care nu sunt potriviți noii populații.

Această situație poate fi identificată observând, pentru persoane din populații diferite care au același scor total la test, dacă există itemi la care răspunsurile sunt diferite. Studiul statistic se bazează în acest caz pe analiza de varianță, variabila dependentă fiind scorul itemului, iar variabilele independente, populația și scorul total al testului.

După efectuarea unei cercetări comparative, pentru identificarea itemilor la care răspunsurile se deosebesc între populații și depistarea cauzelor, se va decide cum trebuie procedat pentru a putea utiliza testul și în noua populație.

În anul 1993 a fost format un comitet internațional (International Test Commission), compus din specialiști din diverse domenii ale psihologiei, care, în 1996, au redactat un set de reguli pentru traducerea testelor (Van de Vijver, Hambleton, 1996).

CAPITOLUL 13 INTERPRETAREA SCORURILOR TESTELOR PSIHOLOGICE

13.1. PRINCIPII GENERALE

Scorul obținut de o persoană la un test, indiferent de formula utilizată pentru calculul acestuia (însușirea scorurilor itemilor, ponderate sau nu, aplicarea formulei de corecție pentru șansa succesului etc.), se numește cota brută.

Valoarea sa nu aduce nici o informație despre priceperile sau cunoștințele subiectului. Faptul că un elev a obținut 60 de puncte la un test de atenție concentrată nu ne spune dacă a realizat o performanță bună, mediocră sau slabă. O cotă brută dobândește semnificație doar dacă ea este raportată:

- la baremuri fixate

sau

- la scorurile realizate de alți subiecți

sau

- la cota brută obținută de aceeași persoană la o testare anterioară.

Prima situație este întâlnită mai ales la testele centrate pe criteriu și la cele folosite în selecție. Uneori, baremul la care se face raportarea rezultatelor unui test este reprezentat de o singură valoare de scor. Aceasta poate fi un scor de secționare (cutoff score) sau un scor critic (critical score) (Truxillo, Donahue, Sulzer, 1996).

Scorul de secționare este un prag care servește la împărțirea subiecților în două categorii: "acceptați" (cei care au un scor superior scorului de secționare) și "respinși" (celelalte persoane). Uneori, scorul de secționare este stabilit de specialiști, care apreciază nivelul minim de cunoștințe, deprinderi ori priceperi necesare pentru a obține succes într-o profesie, la un curs de calificare sau specializare etc. De exemplu, la un test de cunoștințe, ei decid numărul minim de întrebări la care o persoană trebuie să răspundă corect pentru a fi declarată "admisă". Alteleori, scorul de secționare se fixează în urma administrării testului la două grupe contrastante de persoane, astfel încât majoritatea scorurilor obținute de grupa "bună" să se situeze într-o parte a sa, iar majoritatea scorurilor din grupa "slabă" în cealaltă parte. Se recurge la fixarea unui scor de secționare nu numai în cazul selecției profesionale, pentru a elimina o parte din candidați, ci și la alte examinări psihologice, pentru a le oferi un feedback subiecților, care, de multe ori, sunt mai puțin interesați de locul pe care îl ocupă

în clasamentul întocmit pe baza rezultatelor testării, decât de faptul că "au trecut" sau "au căzut" la test (Truxillo, Donahue, Sulzer, 1996).

Scorul critic servește la diferențierea persoanelor care au succes de cele care au insucces într-un anumit domeniu. De obicei stabilirea acestui scor utilizează ecuația de regresie liniară dintre criteriu și predictor, el având rolul de a-i identifica pe subiecții care au valori mari la criteriu, pe baza scorurilor obținute la test. Celelalte două modalități de interpretare a scorului se întâlnesc la majoritatea testelor normative.

Dar, indiferent de felul în care se procedează pentru a desprinde concluzii din rezultatele unei testări psihologice, este bine să se țină seama de următoarele recomandări (Gronlund, 1967):

A. Înainte de a interpreta scorurile obținute la un test se va studia manualul testului, pentru a cunoaște exact semnificația variabilei măsurate sau prezise de test. Simpla denumire a testului nu aduce suficiente informații pentru o interpretare corectă. Un "test de aritmetică", de exemplu, poate conține exerciții simple de calcul sau probleme de raționament.

B. Nu se pot face deducții corecte numai pe baza unui scor de test. Scorul reflectă, pe lângă mărimea variabilei măsurate de test, și aptitudinile persoanei, influențele exercitate asupra ei de mediul cultural din care face parte și de programul instructiv-educativ urmat, adaptarea emoțională, starea de sănătate etc. Un scor scăzut poate fi cauzat, de pildă, de neînțelegerea termenilor utilizați de test sau de lipsa motivației.

C. Când se interpretează rezultatele unei testări psihologice se va avea în vedere la ce vor servi concluziile la care se va ajunge. De exemplu, un scor la un test de inteligență are semnificație diferită, după cum se dorește să se selecteze elevi pentru a forma o clasă de copii capabili de performanțe superioare ori se urmărește să se decidă dacă subiectul trebuie îndrumat spre liceu sau spre o școală profesională.

D. Este preferabil ca interpretarea rezultatelor la un test să se bazeze nu pe scorurile observate, ci pe intervalele de încredere ale scorurilor reale. Această recomandare figurează și în *Standards for Educational and Psychological Tests* din 1974.

E. Orice scor obținut la un test este afectat, mai mult sau mai puțin, și de erori de măsurare a căror mărime nu poate fi determinată. Din acest motiv, se va evita să se ia decizii importante pe baza unui singur scor de test.

13.2. INTERPRETAREA SCORURILOR LA TESTELE DE CUNOȘTINȚE INFORMATIVE

Să considerăm, de exemplu, un test de cunoștințe realizat și administrat de un profesor pentru o clasă, compus din 50 de itemi care conțin câte un exercițiu de calcul aritmetic. Pentru fiecare exercițiu rezolvat corect se acordă un punct, iar cota brută la test a unei persoane este dată de suma cotelor itemilor, deci de numărul de exerciții rezolvate corect.

Pentru a interpreta rezultatul unui elev, profesorul poate proceda în două moduri:

- să calculeze procentul de itemi rezolvați corect de elev din totalul itemilor testului;
- să compare cota brută a elevului cu scorurile celorlalte persoane, pentru a vedea dacă elevul respectiv a obținut un rezultat dintre cele mai bune, unul mediocru sau unul dintre cele mai slabe.

Prima metodă este utilă doar în cazul testelor de cunoștințe cu criterii absolute, când procentul de rezolvări corecte arată cât de aproape se află fiecare elev de performanța minimă necesară. De exemplu, dacă se consideră că pentru promovarea clasei întâi elevii trebuie să fie în stare să rezolve corect cel puțin 45 de exerciții din test, atunci performanța minimă necesară este de $45/50 \cdot 100\% = 90\%$, iar un elev care a rezolvat corect doar 20 de exerciții, adică $20/50 \cdot 100\% = 40\%$, se află cu mult sub nivelul necesar promovării.

Trebuie menționat, însă, că fixarea și utilizarea criteriilor absolute pentru a se lua decizii importante asupra celor examinați (de exemplu, promovarea clasei) impune o verificare foarte atentă a testului din punctul de vedere al validității, fidelității și adecvării la lotul de subiecți, ceea ce nu se prea întâmplă în cazul testelor de cunoștințe informative.

Cota brută a unui elev la un test de cunoștințe elaborat de profesor aduce mai multe informații asupra individului respectiv dacă este comparată cu rezultatele altor persoane din aceeași clasă. Interpretând scorul unui elev prin raportare la poziția sa în cadrul clasei se elimină erorile generate de neadecvarea testului la nivelul de cunoștințe al elevilor examinați (un test prea dificil pentru o clasă furnizează doar scoruri mici, astfel încât cel mai bun rezultat obținut poate fi situat sub ceea ce profesorul consideră că ar fi "nivelul minim necesar de performanță").

13.3. INTERPRETAREA SCORURILOR LA TESTELE CENTRATE PE CRITERIU

Aceste teste folosesc, pentru interpretarea scorurilor, un *criteriu de performanță* fixat înainte de utilizarea lor și care este independent de rezultatele pe care le obțin la test diverse populații.

Stabilirea criteriului se face de către grupul de specialiști care a elaborat testul, ținând cont de comportamentele așteptate din partea persoanelor examinate și de nivelul la care este necesar să se situeze rezultatele, pentru un individ sau pentru un grup de persoane.

De exemplu, în cazul unor teste de cunoștințe, criteriile pot fi de următoarea formă: "Elevul trebuie să traducă cel puțin 80% din cuvintele cuprinse în test." sau "90% din grup trebuie să rezolve corect cel puțin 6 probleme din 10."

Utilizarea testelor centrate pe criteriu în programele instructiv-educative are ca principal scop clasificarea elevilor pe "niveluri" de stăpânire a cunoștințelor, pentru fiecare din obiectivele urmărite de test (Swaminathan, Hambleton, Algina, 1974). Pe baza rezultatelor obținute la test se decide dacă elevul poate trece la programul de instruire următor ori rămâne la programul actual, temporar sau un timp nedefinit.

Deci, pentru un test centrat pe criteriu sunt fixate de către specialiști două sau mai multe categorii ordonate, iar scorurile obținute la test servesc la repartizarea subiecților în aceste categorii. Problema care trebuie rezolvată în acest caz constă în definirea categoriilor.

Să considerăm un test centrat pe criteriu, având m scoruri posibile: $s_1 < s_2 < \dots < s_m$. Se intenționează ca acesta să fie folosit la clasificarea subiecților în k categorii ordonate ($2 \leq k < m$), prima categorie corespunzând celor mai mici scoruri, iar ultima, celor mai mari.

Definirea categoriilor se face prin fixarea a $k+1$ praguri:

$$t_0 < t_1 < \dots < t_k \quad (13.3.1)$$

care realizează o împărțire a intervalului $[s_1, s_m]$ în k intervale disjuncte. Pentru ca orice scor de test să își poată găsi locul într-o categorie, se impun următoarele condiții:

- primul prag să fie mai mic decât cel mai mic scor posibil, iar următorul prag să fie mai mare decât acest scor:

$$t_0 < s_1 < t_1 \quad (13.3.2)$$

- ultimul prag să coincidă cu cel mai mare scor posibil:

$$t_k = s_m \quad (13.3.3)$$

În cea de-a j -a categorie ($j=1, 2, \dots, k$) vor fi repartizate toate persoanele al căror scor aparține intervalului $(t_{j-1}, t_j]$.

Pragurile t_0, t_1, \dots, t_k pot fi fixate de specialiști sau pot fi determinate astfel încât clasificarea definită cu ajutorul lor să semene cât mai mult cu o clasificare existentă, în același număr de clase, realizată de specialiști. Pentru cea de-a doua variantă, se poate utiliza următorul algoritm:

(1) Se administrează testul unei populații repartizate deja în k categorii. De exemplu, în cazul $k=2$, prima categorie poate cuprinde persoane apreciate de specialiști ca având un nivel de cunoștințe "nesatisfăcător", iar a doua categorie, persoane cu un nivel "satisfăcător" al cunoștințelor.

(2) Pentru fiecare posibilitate de fixare a pragurilor t_0, t_1, \dots, t_k , care îndeplinește condițiile (13.3.1), (13.3.2) și (13.3.3):

- Se face clasificarea persoanelor, pe baza scorurilor obținute la test, astfel: un individ care are scorul $x \in (t_{j-1}, t_j]$ este repartizat în categoria j , $j=1, 2, \dots, k$.
- Se calculează coeficientul lui Cohen, K (a se vedea Anexa I), care evaluează concordanța dintre clasificarea existentă și cea corespunzătoare pragurilor t_0, t_1, \dots, t_k . Coeficientul K ia valori între -1 și $+1$, fiind cu atât mai apropiat de $+1$, cu cât cele două clasificări se aseamănă mai mult între ele. Pentru fiecare alegere a pragurilor t_0, t_1, \dots, t_k se obține o altă valoare a coeficientului K .

(3) Se aleg drept margini ale celor k categorii pragurile t_0, t_1, \dots, t_k pentru care coeficientul K este maxim. Acestea realizează o clasificare care se aseamănă cel mai mult posibil cu cea existentă deja în populație.

În cazul în care testul împarte persoanele examinate în două clase ($k=2$), pragul t_1 se numește *nivel acceptabil de performanță* (Guilbert, 1987). Se consideră că un subiect îndeplinește cerințele criteriului dacă a realizat un scor mai mare decât t_1 .

În literatura de specialitate există mai multe păreri privind modul de stabilire a nivelului acceptabil de performanță. Se poate aplica algoritmul descris mai sus sau se poate alege drept barem pragul t_1 pentru care frecvența relativă a persoanelor repartizate în aceeași categorie în ambele clasificări, p_0 , este maximă (Crehan, 1974). Pentru calculul frecvenței p_0 se utilizează formula:

$$p_0 = \frac{\sum_{i=1}^k n_{ii}}{N} \quad (13.3.4)$$

unde:

n_i = numărul de persoane aflate în categoria i și repartizate pe baza scorului obținut la test tot în categoria i , $i=1,2,\dots,k$;
 N = numărul total de persoane cărora li s-a administrat testul.

13.4. INTERPRETAREA SCORURILOR LA TESTELE NORMATIVE

13.4.1. Normele testelor

În cazul testelor normative, interpretarea scorurilor individuale se realizează cu ajutorul unor scări valorice, denumite **norme**, formate în cadrul operației de etalonare a testului.

Prin etalonare se înțelege stabilirea unei scări care să permită determinarea locului ocupat de rezultatele unui subiect față de rezultatele unei populații de referință, suficient de numeroasă, formată din persoane comparabile cu cea examinată.

Obiectivul principal al etalonării îl constituie determinarea repartiției de frecvențe a scorurilor testului într-un eșantion extras din această populație.

Normele testului descriu performanța realizată de eșantionul de indivizi ales. Ele sunt valori ale unei anumite caracteristici a repartiției scorurilor în eșantion. Caracteristica urmărită poate fi:

- frecvența scorurilor mai mici sau mai mari decât o valoare dată sau
- valoarea care împarte șirul scorurilor ordonate în două intervale ale căror frecvențe sunt într-un raport fixat sau
- media scorurilor.

Cu ajutorul normelor se determină modul în care se transformă cotele brute ale testului în cote ușor de interpretat, numite **cote transformate**. Acestea permit compararea rezultatelor între persoane diferite sau, pentru aceeași persoană, între teste diferite.

Pentru a fi cu adevărat utile, normele trebuie să fie descrise detaliat în manualul testului, indicând:

- a. compoziția populației pentru care au fost stabilite;
- b. modul de formare a eșantionului din această populație;
- c. numărul persoanelor din eșantion;

d. caracteristicile eșantionului: sex, vârstă, rasă, statut socio-economic, localizare geografică, nivel de școlarizare etc.

e. gradul în care eșantionul ales este reprezentativ pentru populația specificată;

f. natura și uniformitatea condițiilor în care a fost administrat testul;

g. data testării.

Tabelele de norme incluse în manualele testelor cuprind în paralel cotele brute și cotele transformate corespunzătoare, pentru una sau mai multe populații. Ele oferă informații numai asupra rezultatelor "obișnuite" sau "tipice" pentru populațiile respective. Nu este vorba despre rezultate "ideale", care ar trebui egalate.

Tabelele de norme ajută să se dea răspuns la următoarele întrebări:

- Cum este rezultatul la test al unei persoane în comparație cu cel al altor persoane din aceeași populație?
- Cum este rezultatul unei persoane la un test comparativ cu rezultatul la un alt test?
- Cum este rezultatul la un test comparativ cu rezultatul aceleiași persoane, la același test, administrat cu câțva timp înainte?

Pentru ca interpretarea unui scor individual să fie corectă trebuie ca normele la care se face raportarea acestuia să fie alese în conformitate cu scopul testării. De exemplu, subiectul va fi comparat cu un eșantion provenit din populația din care el face parte, dacă examinarea psihologică are scop diagnostic, sau cu un eșantion extras din populația din care ar urma să facă parte, dacă testul se folosește în cadrul selecției profesionale.

13.4.2. Etalonarea testelor

Etalonarea este una dintre etapele acțiunii de standardizare a testului. Ea se desfășoară în următorii pași:

- (1) Se definește populația pentru care se etalonează testul.
- (2) Se extrage un eșantion din această populație.
- (3) Se administrează testul acestui eșantion.
- (4) Se determină formula de obținere a cotelor transformate din scorurile testului.

În continuare vom detalia conținutul celor patru pași.

- (1) Când se construiește un test, este important să se decidă cine sunt persoanele care vor fi examinate cu el, ce caracteristici are cel mai mare grup din care vor proveni ele. Acesta reprezintă *populația* căreia îi este destinat testul.

Definirea populației se poate face fie concret, prin enumerarea tuturor membrilor ei, fie operațional, prin indicarea unui set de reguli care vor permite să se cunoască, pentru orice persoană, dacă aparține sau nu populației.

De exemplu, populația poate fi formată din "toate persoanele care au vârsta cuprinsă între 14 ani împliniți și 18 ani neîmpliniți, care locuiesc în orașul București și sunt elevi la liceu".

Definiția populației pentru care se face etalonarea unui test este foarte arbitrară: toți indivizii sau numai un lot cu același nivel școlar sau de aceeași vârstă, dintr-un oraș, o regiune sau o țară.

(2) Din populație se va extrage un eșantion de indivizi -numit *grup de normare*-cărui i se va administra testul, pentru a stabili normele.

Este important ca în eșantionul format testul să se comporte la fel ca în întreaga populație: să măsoare sau să prezică aceeași variabilă, iar repartiția scorurilor la test în eșantion să fie asemănătoare celei care s-ar obține dacă s-ar dispune de scorurile întregii populații.

Pentru aceasta, eșantionul constituit trebuie să îndeplinească două cerințe:

- să fie reprezentativ pentru populația din care a fost extras (structura sa să fie asemănătoare structurii populației, în privința zonelor geografice și a mediului de domiciliu, a nivelului socio-economic, a vârstei și a altor caracteristici pe care autorul testului le consideră importante și care ar putea influența rezultatele testului);

- să fie destul de mare, așa încât rezultatele obținute la test să reflecte repartiția acestor valori în populația din care provine.

Obținerea unui eșantion se poate face prin selecție aleatoare simplă sau prin selecție aleatoare stratificată.

În primul caz, fiecare membru al populației are șanse egale de a intra în componența eșantionului. Dacă eșantionul este numeros, atunci în el se vor găsi persoane aparținând tuturor zonelor geografice, claselor sociale, naționalităților prezente în populație, cu aceeași frecvență relativă ca în populație, deci eșantionul va fi reprezentativ pentru populație.

Atunci când volumul eșantionului este mic, pentru obținerea unui eșantion reprezentativ este necesar să se facă o selecție aleatoare stratificată. În acest scop, se determină variabilele de mediu (sex, vârstă, rasă, clasă socială, nivel educațional etc.) care au o influență mare asupra rezultatelor aplicării testului. Se face apoi o clasificare a populației după valorile acestor variabile și se stabilește care trebuie să fie compoziția eșantionului astfel încât fiecare clasă să fie reprezentată în eșantion proporțional cu frecvența ei relativă în populație.

Exemplu.

Dacă populația este formată din 100000 de persoane și se formează clase în funcție de sex și mediul de domiciliu, conform tabelului următor:

Clasa	Sex	Mediu de domiciliu	Frecvența	Frecvența relativă
C1	bărbați	urban	27000	0.27
C2	bărbați	rural	22000	0.22
C3	femei	urban	23000	0.23
C4	femei	rural	28000	0.28
Total			100000	1.00

atunci într-un eșantion compus din 2000 de persoane frecvențele claselor vor fi:

Clasa	Frecvența relativă	Frecvența
C1	0.27	540
C2	0.22	440
C3	0.23	460
C4	0.28	560
Total	1.00	2000

Din fiecare clasă se extrag, prin selecție aleatoare simplă, persoanele care vor intra în componența eșantionului.

Principala dificultate în cazul selecției aleatoare stratificate constă în determinarea variabilelor după care se face împărțirea în clase. Se recomandă (Kline, 1993) ca ele să se aleagă dintre variabilele care corelează semnificativ cu scorurile testului. Numărul lor nu va fi prea mare, pentru a nu se forma prea multe clase, ceea ce ar necesita alcătuirea unui eșantion foarte numeros. În general nu se aleg mai mult de patru variabile. Frecvent, variabilele după care se face clasificarea sunt sexul, vârsta și clasa socială.

Volumul eșantionului se stabilește astfel încât din fiecare clasă să se extragă cel puțin 300 de persoane (Kline, 1993). Se deduce de aici că populațiile generale, care sunt eterogene și din acest motiv necesită utilizarea mai multor variabile pentru formarea claselor, impun obținerea unor eșantioane foarte numeroase. Pentru a se face economie de timp și de bani, adesea se renunță la

stabilirea normelor pentru o populație generală și se utilizează populații mai puțin numeroase și mai omogene, pentru care eșantioanele pot fi de volum mai redus. În acest fel se obțin *norme locale ale testului*.

Dacă se constată că un subgrup al eșantionului, care poate fi identificat printr-o valoare a unei variabile de mediu (de exemplu "femei" sau "persoane din mediul urban") realizează la test scoruri mult mai mari sau mult mai mici decât restul eșantionului, se vor determina *norme ale subgrupului* și acestea vor fi folosite în practică atunci când persoanele al căror scor trebuie interpretat au aceeași valoare a variabilei de mediu ca și subgrupul.

(3) Dacă eșantionul are un volum foarte mare ori este format din persoane care domiciliază sau lucrează în locuri diferite, testul nu poate fi administrat deodată tuturor indivizilor. Examinarea cu ajutorul testului se va face pe subgrupe. Dar, pentru ca rezultatele să nu fie distorsionate, condițiile de administrare trebuie să fie identice pentru toate subgrupele.

(4) În urma analizei scorurilor realizate la test de persoanele din eșantion se stabilesc normele testului. Apoi se determină modul în care se obțin cotele transformate din cotele brute. În acest scop se folosesc mai frecvent două tipuri de transformări:

- *transformări liniare*, care schimbă doar media și abaterea standard a rezultatelor, păstrând nemodificată repartiția cotelor, deci ordinea indivizilor din eșantion. Prin astfel de transformări se obțin cotele standard z și cotele standardizate;

- *transformări neliniare*, sau *transformări de arie*, care schimbă repartiția cotelor brute, făcând-o să semene cu repartiția uniformă (în cazul centilelor, decilelor sau al cuartilelor) ori cu cea normală (în cazul când se urmărește obținerea unei scale normalizate cu 5, 7 sau 9 trepte). Acest fel de transformări schimbă unitatea scalei, astfel încât frecvența relativă a cotelor brute care ocupă un interval de valori particular va fi egală cu aria suprafeței cuprinsă între intervalul transformat, desenat pe axa absciselor, și:

- curba repartiției uniforme, pentru rangurile centile, decile și cuartile sau

- curba repartiției normale, pentru clasele scării normalizate.

Cotele transformate rezultate prin transformări liniare sau neliniare indică poziția relativă a subiectului într-o populație.

Un alt tip de transformări utilizate în etalonare sunt cele care conduc la *clase echivalente* sau la *vârste echivalente*. Ele se aplică pentru teste care măsoară variabile fiziologice sau psihologice ale căror valori cresc cu vârsta (de exemplu înălțimea, greutatea, bogăția vocabularului, performanța la citire etc.).

Aceste cote transformate descriu performanța subiectului sub forma indicării unei populații a cărei mediană sau medie a cotelor brute este aproximativ egală cu cota brută a persoanei respective. Unele teste de inteligență utilizează pentru interpretarea scorurilor, în locul vârstelor echivalente, normele de vârstă mintală.

13.4.3. Clasele echivalente

Aceste cote transformate sunt folosite frecvent pentru testele de cunoștințe standardizate, destinate claselor primare.

Tabelele cu clasele echivalente conțin, pentru fiecare lună de școală a fiecărei clase, mediana cotelor brute la test. Clasa echivalentă este exprimată printr-un număr zecimal: partea întregă reprezintă clasa, iar cifra zecimală, luna de școală. Se face presupunerea că în cele două luni de vacanță de vară cotele la test ale elevilor nu se modifică.

Pentru interpretarea rezultatului obținut de un subiect se caută în tabelul de norme clasa echivalentă corespunzătoare cotei brute respective.

Exemplu.

În tabelul 13.4.3.1. sunt notate clasele echivalente pentru un test de aritmetică ale cărui cote brute iau valori între 0 și 100, destinat elevilor din clasele a II-a, a III-a și a IV-a.

Tabelul 13.4.3.1
Clasele echivalente pentru un test de aritmetică

Clasa	Cota brută	Clasa	Cota brută	Clasa	Cota brută
2.0	15	3.0	25	4.0	55
2.1	15	3.1	27	4.1	58
2.2	16	3.2	28	4.2	62
2.3	17	3.3	30	4.3	64
2.4	17	3.4	32	4.4	66
2.5	18	3.5	36	4.5	67
2.6	19	3.6	39	4.6	68
2.7	20	3.7	44	4.7	69
2.8	22	3.8	47	4.8	69
2.9	23	3.9	51	4.9	70

Un elev care în prima săptămână de școală din clasa a IV-a a obținut la test o cotă brută egală cu 27 are clasa echivalentă 3.1, deci este rămas în urmă la aritmetică cu aproape un an.

Popularitatea utilizării claselor echivalente este dată de faptul că furnizează cote transformate ușor de înțeles și permite compararea rezultatelor obținute la mai multe teste.

Exemplu.

Un copil, care este în a doua lună de școală din clasa a patra, a obținut la trei teste următoarele clase echivalente:

Limba română : 4.5

Aritmetică : 3.0

Istorie : 5.4

Analizând aceste rezultate, părinții și profesorii vor înțelege că performanțele la Limba română sunt apropiate de cele ale celorlalți elevi în această perioadă a anului școlar, în timp ce scorul de la testul de Aritmetică denotă o rămânere în urmă de peste un an, iar cel de la testul de Istorie indică o depășire a cunoștințelor cu peste un an.

Trebuie subliniat că valoarea cotei brute indicată în tabelul de norme pentru fiecare clasă echivalentă nu reprezintă o performanță care este necesar să fie atinsă. Ea este doar un rezultat tipic al unui grup de elevi cu același nivel de școlarizare. Aproximativ 50% dintre persoanele grupului respectiv au rezultate mai bune și aproximativ 50% au rezultate mai slabe decât această cotă.

Principala deficiență a claselor echivalente constă în faptul că aceleași "distanțe" între clase echivalente îi pot corespunde distanțe diferite între cotele brute ale testului. De exemplu, în tabelul 13.4.3.1, unei diferențe de un an îi corespunde o distanță de 10 puncte, dacă se compară clasele echivalente 2.0 și 3.0, sau de 30 de puncte, dacă se compară clasele echivalente 3.0 și 4.0. Deci, informația "Elevul este în urmă cu un an la aritmetică" are semnificație diferită, în funcție de momentul în care este emisă.

Explicația constă în faptul că dezvoltarea intelectuală și acumularea de cunoștințe sunt procese care nu au o evoluție liniară în timp. În perioadele școlare în care acumularea de cunoștințe este masivă, unei diferențe mici de clase echivalente (una sau două luni școlare) îi corespunde o diferență mare de cote brute la test. În plus, curba dezvoltării intelectuale relative la vârstă diferă de la o persoană la alta.

Din aceste motive se recomandă să se folosească transformarea cotelor brute în clase echivalente numai pentru aritmetică, citire și compunere și doar

în perioada claselor primare, când procesul de instruire este asemănător în toate școlile.

Un procedeu care poate fi aplicat pentru obținerea tabelelor cu clase echivalente este următorul (Flanagan, 1955):

(1) Se administrează testul, în același timp, unor loturi mari de elevi din clase consecutive, deci cu "distanță" de o clasă echivalentă între ele.

(2) Pentru fiecare clasă se determină mediana cotelor brute.

(3) Se interpolează punctele consecutive (clasă, mediană), fie liniar, fie printr-o funcție crescătoare.

(4) Pentru fiecare clasă echivalentă aflată între clasele pentru care s-a determinat mediana, din lună în lună, se citește de pe graficul funcției interpolatoare cota brută a testului și se rotunjește la cel mai apropiat întreg.

Este important ca loturile de elevi utilizate la stabilirea claselor echivalente să fie numeroase și să conțină persoane cu niveluri școlare diferite (atât avansați, cât și retardați).

13.4.4. Vârstele echivalente

Aceste cote transformate indică, pentru fiecare vârstă cronologică exprimată sub forma unei perechi de numere (an, lună), mediana sau media (după preferința celui care a făcut etalonarea) cotelor brute la test. Ele sunt folosite mai ales pentru teste de inteligență (tabelele de norme conțin "vârsta mintală" echivalentă) și pentru teste de cunoștințe (tabelele de norme conțin "vârsta la aritmetică" sau "vârsta la citire" etc.). La fel ca și clasele echivalente, aceste cote transformate sunt potrivite pentru nivelul școlar elementar, când dezvoltarea mintală și influența educației sunt continue și nu oscilează.

Vârstele echivalente au aceleași avantaje și deficiențe ca și clasele echivalente.

13.4.5. Normele de vârstă mintală

Conceptul de *vârstă mintală* a fost introdus în scalele Binet-Simon revizuite în anul 1908. Acestea grupează itemii pe niveluri de vârstă: itemii care sunt rezolvați corect de majoritatea copiilor de 7 ani din eșantionul utilizat la etalonare sunt plasați la nivelul vârstei de 7 ani ș.a.m.d. Scorul unui copil la acest test corespunde celui mai înalt nivel pe care l-a rezolvat cu succes. De pildă, dacă un copil de 10 ani (vârsta cronologică) rezolvă corect itemii aflați la nivelul vârstei de 12 ani, se va spune că are vârsta mintală de 12 ani.

Trebuie menționat că o unitate de vârstă mintală (un an mintal) nu are aceeași semnificație la orice vârstă cronologică. Datorită faptului că dezvoltarea intelectuală este rapidă la vârste fragede și își încetinește ritmul la maturitate, importanța unei diferențe de un an mintal scade cu vârsta. O diferență de un an mintal la vârsta cronologică de 3-4 ani înseamnă tot atât de mult ca o diferență de 3 ani mintali la vârsta cronologică de 9-12 ani (Anastasi, 1954).

Pentru a putea interpreta mai ușor rezultatele testelor de inteligență, fără a mai face referire de fiecare dată la vârsta cronologică, s-a definit *coeficientul de inteligență* (IQ = Intelligence Quotient). Acesta este egal cu raportul dintre vârsta mintală și vârsta cronologică, multiplicat cu 100. Deci, o valoare de 100 a acestui coeficient indică egalitatea vârstei mintale cu vârsta cronologică.

Coeficientul de inteligență permite compararea persoanelor de vârste diferite în privința dezvoltării intelectuale. Dar, el nu poate fi utilizat pentru orice test de inteligență destinat adulților. La un test ca Stanford-Binet, de exemplu, media scorurilor obținute de persoane de aceeași vârstă cronologică, mai mare de 15 ani, nu crește semnificativ de la un an la altul. Deci, indiferent de vârsta lor cronologică, adulții care nu sunt debili mintal vor obține la test, în general, vârsta mintală de 15 ani. Numărătorul coeficientului de inteligență fiind același (egal cu 15), valoarea raportului prin care se calculează acesta va depinde doar de vârsta cronologică, mai precis va descrește cu vârsta.

13.4.6. Cotele standard z și cotele standardizate

Aceste cote transformate servesc la indicarea poziției unei persoane față de o anumită populație.

O cotă standard z arată cu câte unități de abatere standard se distanțează cota brută la test a unui subiect de media cotelor brute în populația cu care se face comparația.

Cotele z se calculează pe baza cotelor brute ale testului, folosind formula:

$$z = \frac{x-m}{\sigma} \quad (13.4.6.1)$$

unde:

z = cota z corespunzătoare scorului la test x ;
 m, σ = media și abaterea standard pentru cotele brute la test în eșantionul folosit la etalonare.

Cotele standard se folosesc pentru a compara performanțele unui individ la două sau mai multe teste, dacă mediile și dispersiile scorurilor acestora au fost calculate pe baza aceluiași eșantion de persoane.

De exemplu, dacă un elev are cotele z la un test de aritmetică și la unul de limba română egale cu 1.8, respectiv cu 0.3, se poate aprecia că elevul este mai bine pregătit la aritmetică decât la limba română, chiar dacă el a obținut aceeași cotă brută la ambele teste.

Cotele standard sunt ușor de interpretat atunci când cotele brute ale testului sunt repartizate aproximativ normal în populația care a servit la etalonare. În acest caz, în intervalul $[-1, +1]$ se află cotele z pentru 68% dintre persoane. Despre un subiect care are o cotă standard $z=0.72$ se poate afirma că a obținut un rezultat "obișnuit", în timp ce o persoană cu o cotă z egală cu 3.6 are un scor "neobișnuit de mare" (mai puțin de 0.13% dintre cei testați cu ocazia etalonării au mai realizat un asemenea scor).

Cotele standardizate au aceleași proprietăți ca și cotele standard z , singura diferență între ele fiind valoarea mediei și cea a abaterii standard (a se vedea Anexa II).

13.4.7. Rangurile centile, decile și cuartile

Rangul centil arată câte persoane s-ar afla în urma unui anumit subiect, dacă acesta ar face parte dintr-un grup cu 100 de indivizi. De exemplu, dacă 85% din cei testați obțin scoruri mai mici decât 60 și 15% realizează scoruri mai mari, atunci scorul 60 va avea rangul centil 85.

Prezentăm în continuare, pe un exemplu, pașii algoritmului de determinare a rangurilor centile, prelucrând rezultatele obținute la un test de un eșantion format din 200 de persoane.

(1) Se ordonează crescător cotele brute ale testului (x) și se notează pentru fiecare valoare numărul de persoane care au realizat scorul respectiv (N). Se obține tabelul 13.4.7.1.

Tabelul 13.4.7.1
Frecvențele cotelor brute la test

x	10	11	12	14	15	16	18	21	22	23
N	10	20	20	30	10	40	30	20	10	10

(2) Se trasează poligonul frecvențelor cumulate (Figura 13.4.7.1), procedînd astfel:

- În jurul fiecărei valori de scor $x_i, i=1,2,\dots,n$, se construiește intervalul $[x_i-0.5, x_i+0.5]$.
- Se calculează valoarea fiecărei ordonate y_{i-} , $i=1,2,\dots,n$, corespunzătoare abscisei $x_i-0.5$, făcînd raportul dintre numărul persoanelor care au scor mai mic decît x_i și numărul total de persoane în eșantion și apoi înmulțind cu 100.
- Se calculează valoarea fiecărei ordonate y_{i+} , $i=1,2,\dots,n$, corespunzătoare abscisei $x_i+0.5$, făcînd raportul dintre numărul persoanelor care au scor mai mic sau egal cu x_i și numărul total de persoane în eșantion și apoi înmulțind cu 100. Se observă că y_{i+} este egal cu $y_{(i+1)-}$.
- Se unesc, pe rînd, punctele $(x_i-0.5, y_{i-}), (x_i+0.5, y_{i+}), (x_{i+1}-0.5, y_{i+1}), \dots, (x_n+0.5, y_{n+1})$, prin segmente de dreaptă.
- În exemplul nostru $n=10$. Valorile x_i, y_{i-} și y_{i+} , $i=1,2,\dots,10$, sunt notate în tabelul 13.4.7.2.

Tabelul 13.4.7.2
Valorile x_i, y_{i-} și y_{i+} utilizate la trasarea poligonului frecvențelor cumulate

i	1	2	3	4	5	6	7	8	9	10
x_i	10	11	12	14	15	16	18	21	22	23
y_{i-}	0	5	15	25	40	45	65	80	90	95
y_{i+}	5	15	25	40	45	65	80	90	95	100

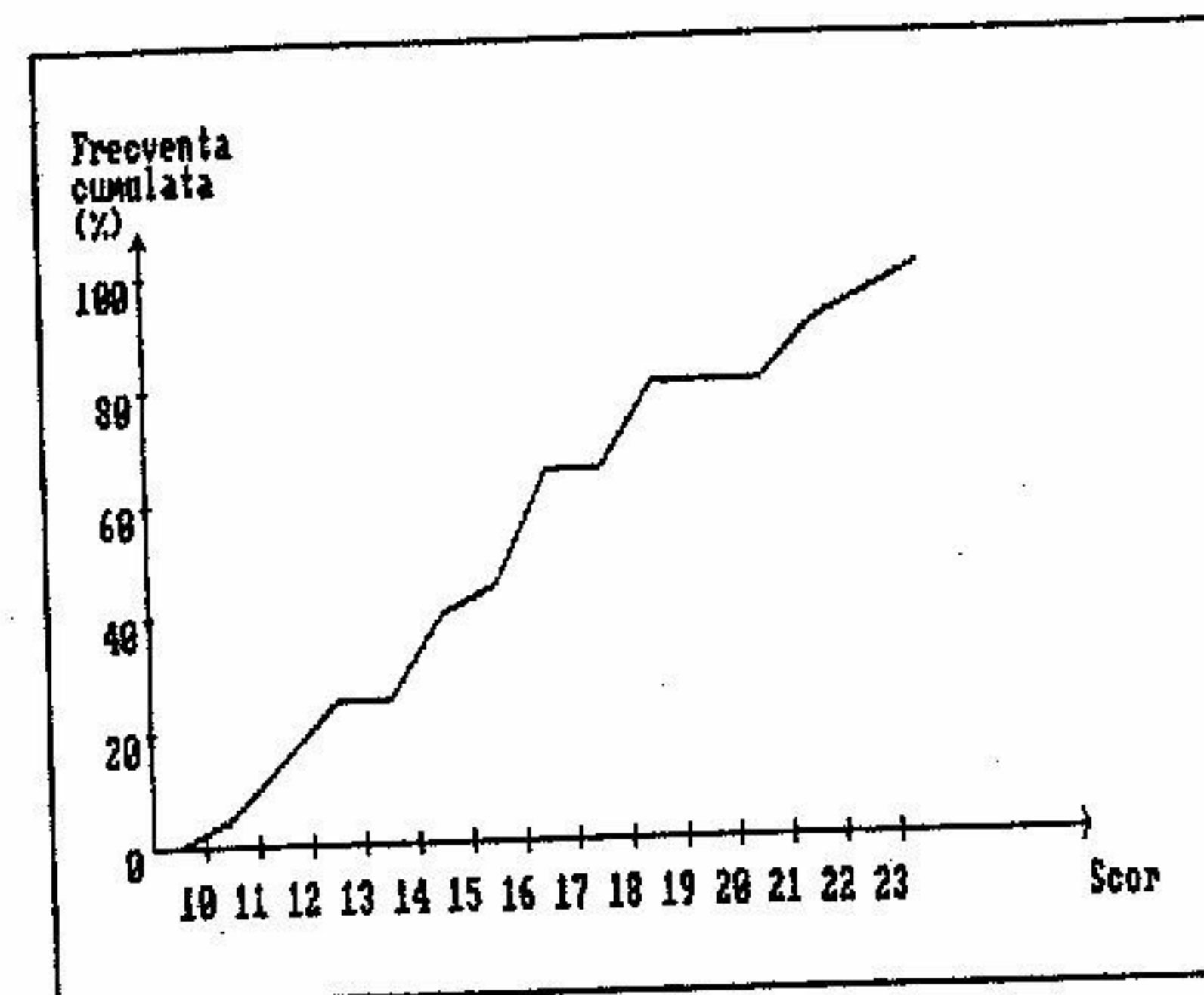


Figura 13.4.7.1
Poligonul frecvențelor cumulate

(3) Pentru a determina rangul centil corespunzător unui scor dat, x , se citește de pe grafic ordonata punctului care are abscisa x . De exemplu, pentru scorul $x=14$:

$$i=4, y_{i-}=25, y_{i+}=40.$$

Rangul centil este valoarea y care corespunde scorului $x=14$ în ecuația dreptei ce trece prin punctele

$$(x_i-0.5, y_{i-}) \text{ și } (x_i+0.5, y_{i+}):$$

$$\frac{y-25}{40-25} = \frac{x-13.5}{14.5-13.5}$$

sau

$$y = 15x - 177.5$$

Deci, scorul 14 are rangul $15 \cdot 14 - 177.5 = 32.5$.

Obişnuit, rangul centil se rotunjește la cel mai apropiat întreg.

(4) Pentru a determina scorul care corespunde unui rang centil dat, y , se citește de pe grafic abscisa punctului care are ordonata y .

De exemplu, rangului centil 30 îi corespunde pe grafic un punct situat pe dreapta care trece prin punctele $(13.5, 25)$ și $(13.5, 40)$.

Înseamnă că două ranguri centile consecutive vor corespunde unor valori de scor foarte apropiate, deci unor persoane foarte asemănătoare - în privința variabilei măsurate de test-, în primul caz, și unor scoruri mai îndepărtate, deci unor indivizi destul de deosebiți unul de altul, în cel de-al doilea.

Atunci când repartiția scorurilor testului este uniformă, deci fiecare cotă brută apare cu aceeași frecvență, unor intervale egale ca lungime pentru cotele brute le corespund intervale de lungimi egale pentru rangurile centile.

Folosirea rangurilor centile își găsește justificarea numai atunci când testul are un număr mare de scoruri posibile. În caz contrar, este de preferat rangul decil sau rangul cuartil. Acestea se obțin într-un mod asemănător rangurilor centile, doar că aria cuprinsă între axa absciselor și curba de repartiție a scorurilor nu se mai împarte în 100 de părți egale, ci în 10, pentru decile, sau în 4, pentru quartile. Centilele care au rangurile 10, 20, ..., 90 coincid cu decilele, iar centilele cu rangurile 25, 50 și 75 sunt egale cu quartilele.

13.4.8. Clasele scării normalizate

La fel ca și rangurile centile, decile și quartile, clasele scării normalizate se obțin prin transformări de arie. Diferă doar curba de repartiție utilizată: în primul caz este repartiția uniformă, în cel de-al doilea, cea normală.

Procedeul de lucru pentru obținerea scării normalizate cu 5, 7 și, respectiv, 9 clase este asemănător. În toate cele trei cazuri se folosește curba de repartiție a legii normale cu media zero și cu abaterea standard σ . Axa reală se decupează în 5, 7 sau, respectiv, în 9 intervale, dispuse câte două simetric față de origine, în felul următor:

- primul și ultimul interval au lungime nelimitată;
- celelalte intervale au lungime finită, egală cu σ , în primul caz, cu 0.66σ în cel de-al doilea caz și cu 0.5σ în cel de-al treilea caz;
- intervalul central conține în mijloc originea axei.

Pe baza tabelelor legii normale se determină, pentru fiecare clasă, aria suprafeței cuprinsă între intervalul corespunzător clasei și curba densității de repartiție a legii normale. Aceasta va indica frecvența relativă a scorurilor pe care le va conține clasa (Figura 13.4.8.1).

Valorile găsite, exprimate procentual, sunt următoarele :

- pentru 5 clase: 7%, 24%, 38%, 24%, 7%;
- pentru 7 clase: 5%, 11%, 21%, 26%, 21%, 11%, 5%;
- pentru 9 clase: 4%, 7%, 12%, 17%, 20%, 17%, 12%, 7%, 4%.

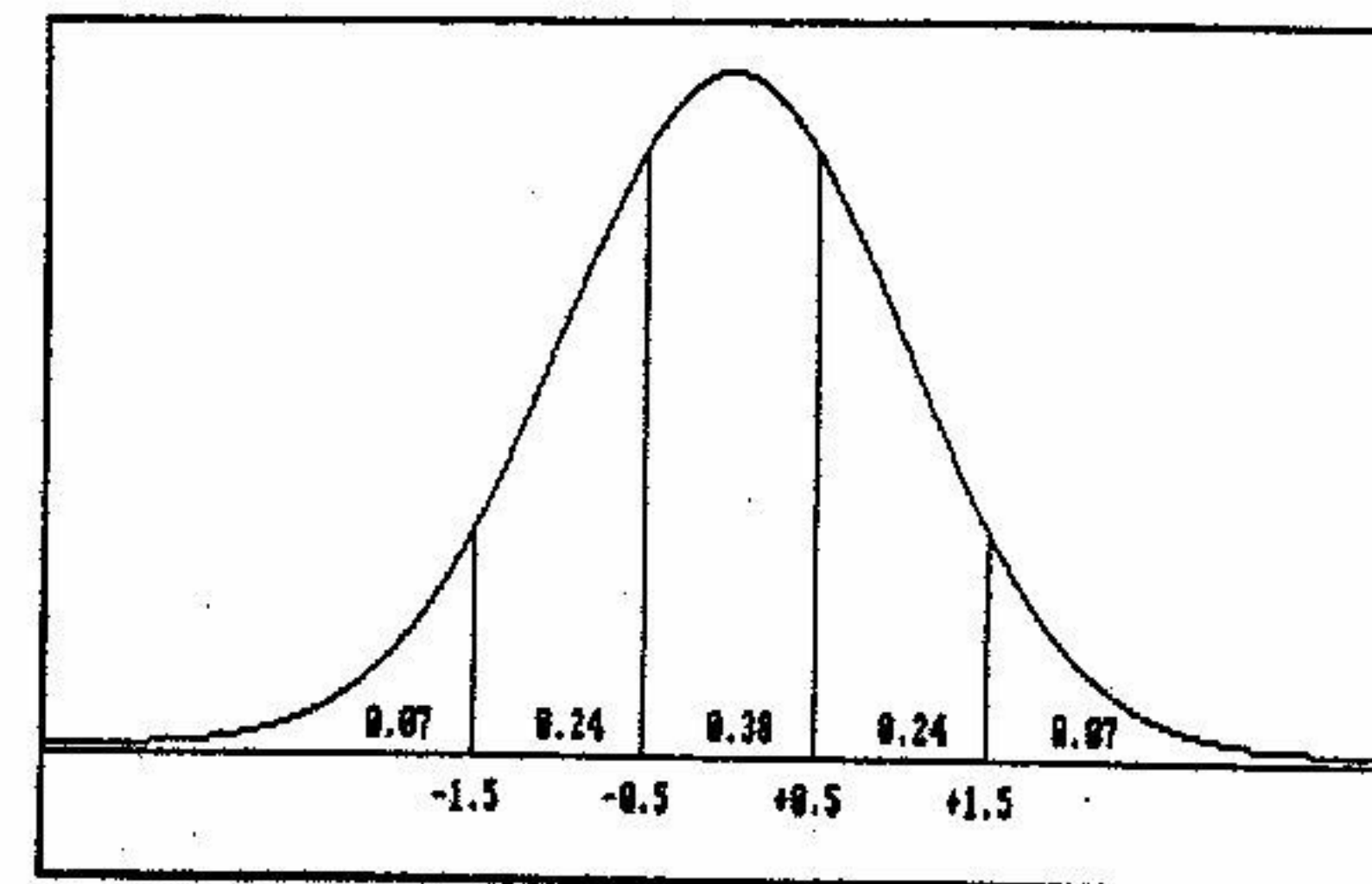


Figura. 13.4.8.1

Împărțirea suprafeței cuprinsă sub curba densității de repartiție a legii normale în cinci clase.

Concret, în cazul scării normalizate cu 5 clase, considerând scorurile ordonate crescător:

- prima clasă va conține primele 7% din scoruri, cele mai mici;
- a doua clasă va conține următoarele 24% din scoruri ș.a.m.d.

Pe baza procentelor se determină, pentru fiecare clasă, limitele intervalului de valori al scorurilor:

- Prima clasă va avea limita inferioară egală cu cel mai mic scor posibil.
- Ultima clasă va avea limita superioară egală cu scorul cel mai mare posibil.
- Celelalte limite vor fi calculate ținând cont de valorile scorurilor cuprinse în fiecare clasă. Se poate proceda astfel :

- Se desenează poligonul frecvențelor cumulate ale scorurilor.
- Se determină punctele graficului corespunzătoare frecvențelor cumulate ale claselor (7%, 31%, 69%, 93% și 100% pentru scara normalizată cu 5 clase).
- Se citesc de pe abscisă valorile scorurilor pentru aceste puncte; ele vor fi limitele intervalelor.

Rezultatele etalonării vor fi prezentate sub forma unui tabel, în care pentru fiecare clasă vor fi indicate limitele intervalului de valori al scorurilor.

Prin transformările descrise, scorurile obținute la test de către eșantionul selecționat vor primi valori din mulțimea

$$\{1, 2, \dots, c\},$$

c reprezentând numărul de clase ($c=5$ sau $c=7$ sau $c=9$). Scorurile transformate vor avea o repartiție aproximativ normală.

Prezentăm algoritmul de transformare a scorurilor în 5 clase normalizate, folosind datele din exemplul precedent:

(1) Notăm într-un tabel frecvențele cotelor brute ale testului (Tabelul 13.4.8.1).

Tabelul 13.4.8.1

Frecvențele absolute și cumulate ale scorurilor testului

Scor (x)	10	11	12	14	15	16	18	21	22	23
Frecvența absolută	10	20	20	30	10	40	30	20	10	10
Frecvența cumulată	10	30	50	80	90	130	160	180	190	200
Frecvența cumulată, în procente (y)	5	15	25	40	45	65	80	90	95	100

(2) Reprezentăm grafic frecvențele cumulate, interpolând liniar între punctele (x,y) consecutive (Figura 13.4.8.2)

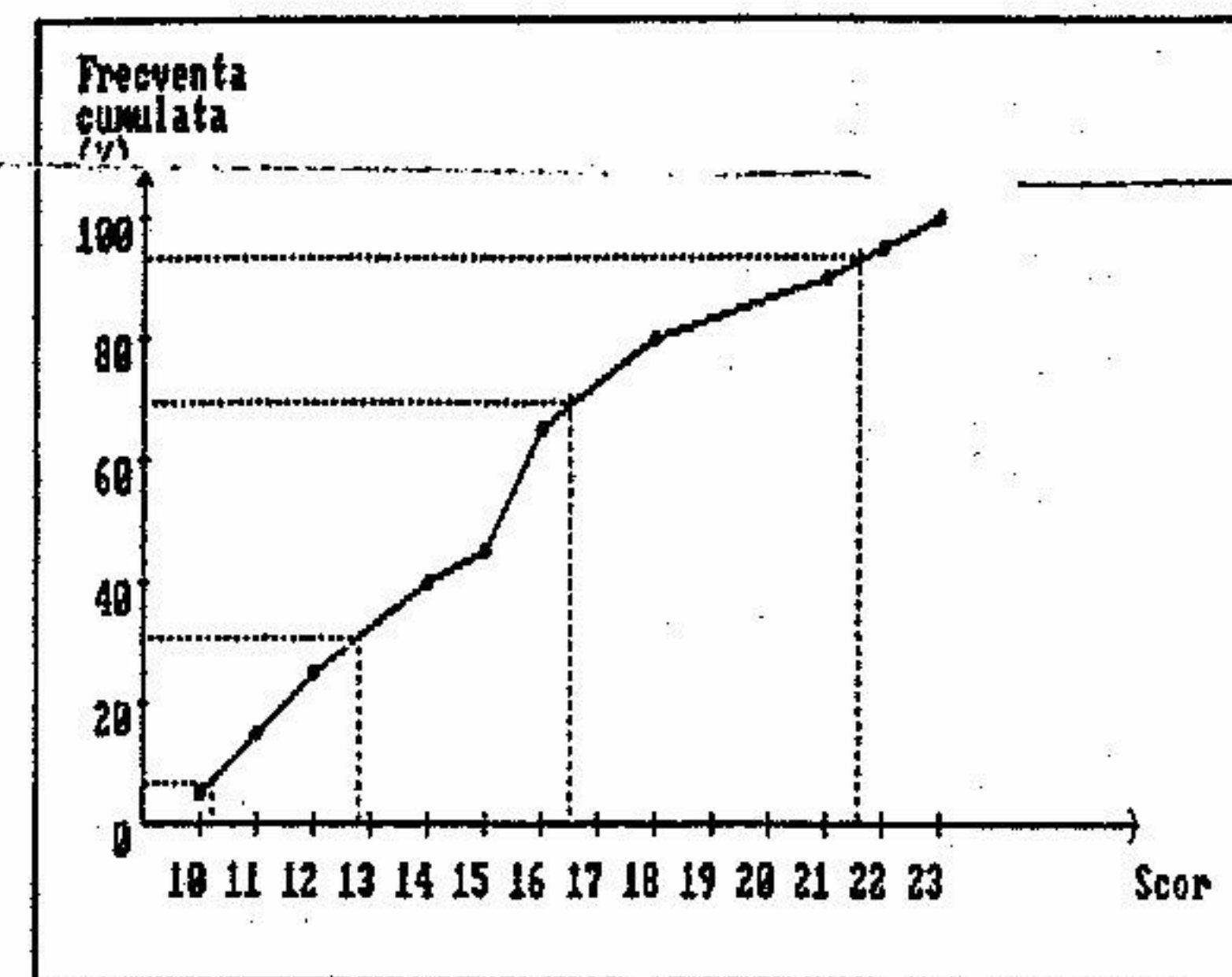


Figura 13.4.8.2

Determinarea intervalelor scorurilor la transformarea în cinci clase normalizate

(3) Prima clasă trebuie să cuprindă 7% din persoane. Punctul de pe grafic care are ordonata egală cu 7 se găsește pe dreapta care trece prin punctele (10, 5) și (11, 15). Ecuația acestei drepte este

$$\frac{y-5}{15-5} = \frac{x-10}{11-10}$$

sau

$$y=10x-95$$

Pentru $y=7$ se obține $x=10.2$. Prima clasă va avea limita inferioară egală cu cea mai mică valoare posibilă de scor și limita superioară egală cu 10.2.

(4) Primele două clase trebuie să cuprindă $7\%+24\%=31\%$ din totalul persoanelor. Punctul de pe grafic care are ordonata egală cu 31 se află pe dreapta care trece prin punctele (12, 25) și (14, 40). Această dreaptă are ecuația :

$$\frac{y-25}{40-25} = \frac{x-12}{14-12}$$

sau

$$y=7.5x-65$$

Pentru $y=31$ se obține $x=12.8$. Deci a doua clasă va conține intervalul (10.2, 12.8].

(5) În același mod vom găsi intervalele pentru celelalte clase (Tabelul 13.4.8.2).

Tabelul 13.4.8.2
Intervalele de scoruri pentru fiecare clasă normalizată

clasa	interval
1	≤ 10.2
2	(10.2, 12.8]
3	(12.8, 16.5]
4	(16.5, 21.6]
5	> 21.6

(6) Utilizând tabelul 13.4.8.2 se determină clasa normalizată care corespunde fiecărei cote brute (Tabelul 13.4.8.3).

Tabelul 13.4.8.3
Correspondența cotă brută - clasă normalizată

cota brută	≤ 10	11	12	13	14	15	16	17	18	19	20	21	≥ 22
clasa normalizată	1	2	2	3	3	3	3	4	4	4	4	4	5

Prin această transformare se schimbă forma repartiției scorurilor (Figura 13.4.8.3).

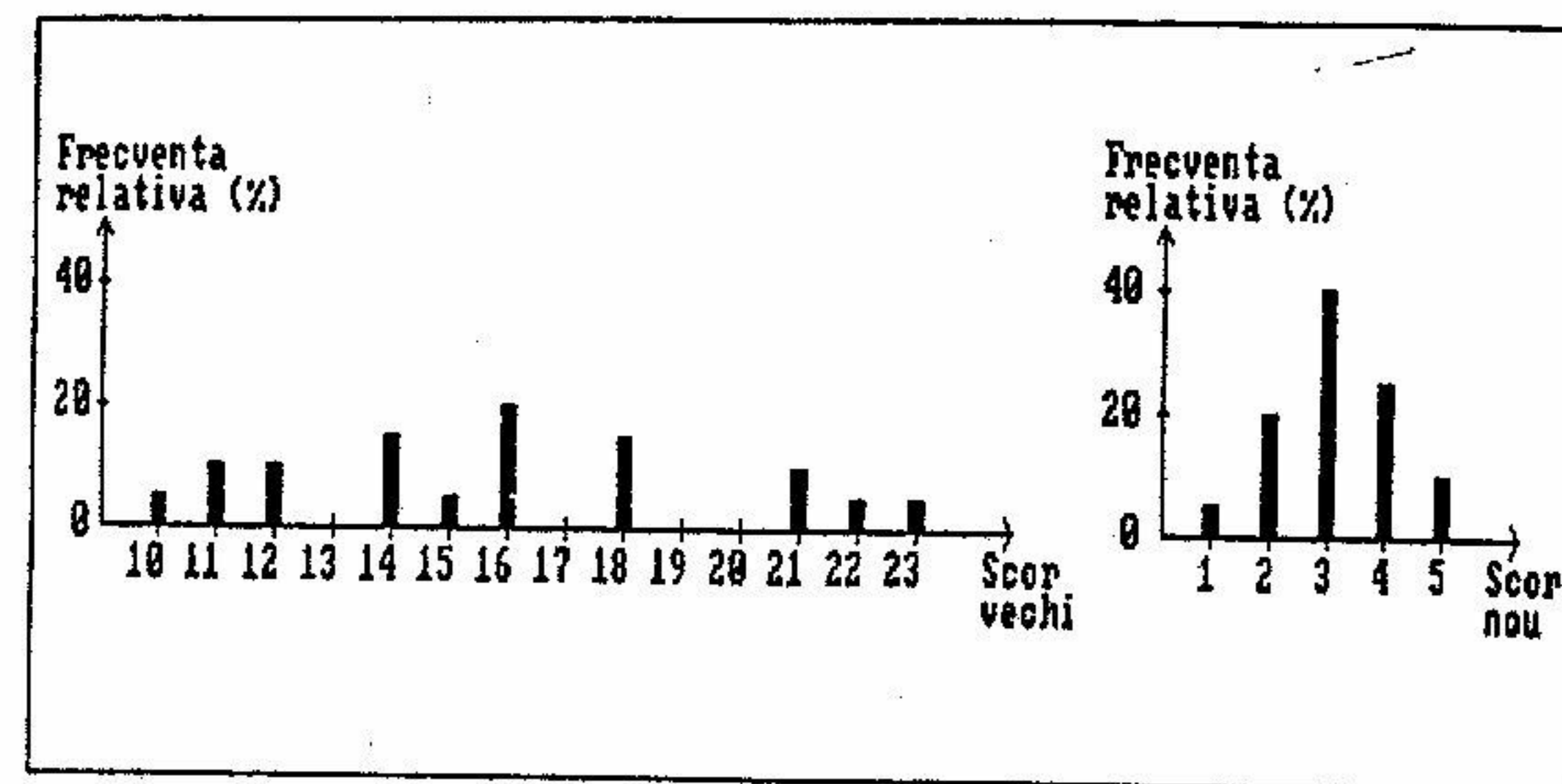


Figura 13.4.8.3
Schimbarea formei repartiției scorurilor prin transformarea în cinci clase normalizate

Există și cazuri când o transformare în clase normalizate nu are sens. De exemplu, atunci când frecvența relativă a primului scor este mai mare decât 11% este imposibil de stabilit limitele claselor 1 și 2 pentru o transformare normalizată în 9 clase.

13.5. INTERPRETAREA SCORURILOR LA CHESTIONARELE DE PERSONALITATE

13.5.1. Descrierea procedurii obișnuit

Interpretarea scorurilor obținute de un subiect la majoritatea chestionarelor de personalitate are la bază adjective și propoziții scurte descriind comportamente care sunt grupate în liste atașate scalelor sau unor grupări de scale.

Prezentăm câteva exemple referitoare la Inventarul Psihologic California (CPI).

Pentru fiecare dintre primele 20 de scale din CPI există două asemenea liste: una corespunzătoare scorurilor mici și alta corespunzătoare scorurilor mari.

De exemplu, scala So (care evaluează gradul de interiorizare a valorilor normative și culturale) are atașate următoarele liste (Pitariu, Albu, 1993):

Scoruri mici: rebel, nesatisfăcut, nonconformist.

Scoruri mari: conștiincios, bine organizat, productiv.

Pentru unele perechi de scale, cum ar fi

Re (responsabilitate) și Cs (capacitate de statut),

Do (dominanță) și Re (responsabilitate),

Do (dominanță) și Sy (sociabilitate)

au fost întocmite câte patru liste, conform următoarei scheme (Megargee, 1972; Pitariu, 1985):

Lista 1:

scala 1: scoruri mici

scala 2: scoruri mari

Lista 2:

scala 1: scoruri mari

scala 2: scoruri mari

Lista 3:

scala 1: scoruri mici

scala 2: scoruri mici

Lista 4:

scala 1: scoruri mari

scala 2: scoruri mici

Pentru ca o caracterizare întocmită cu ajutorul listelor atașate scalelor să concorde cât mai mult cu situația reală trebuie să fie soluționate corect următoarele două probleme:

- Adjectivele sau propozițiile conținute în liste să fie adecvate scalei sau grupărilor de scale.

- Din liste să fie selectate acele adjective sau propoziții care au o probabilitate mare de a fi potrivite persoanei testate.

Algoritmii utilizați pentru rezolvarea acestor probleme în cazul chestionarelor de personalitate utilizate în practică nu sunt totdeauna corecți. Așa este, de exemplu, metoda aplicată la formarea listelor atașate scalelor din CPI (Gough, 1989, 1991; Megargee, 1972). Ea se bazează pe coeficientul de corelație liniară între scorurile scalelor și scorurile adjectivelor (scorul unui subiect la un adjectiv este egal cu frecvența alegerii adjectivului pentru caracterizarea persoanei). În lista corespunzătoare scorurilor mari ale unei scale sunt introduse adjective care corelează pozitiv semnificativ statistic cu scorurile scalelor, deci care sunt mai potrivite persoanelor care obțin scoruri mari decât celor care realizează scoruri mici. Dar, nu există siguranța că aceste adjective sunt într-adevăr potrivite pentru majoritatea celor care au scoruri mari și, de asemenea, nu se poate ști dacă ele nu sunt valabile pentru un procent însemnat din mulțimea subiecților cu scoruri mici.

Faptul că listele sunt atașate "scorurilor mari", respectiv "scorurilor mici", ale scalelor prezintă mai multe dezavantaje și constituie o sursă de erori în interpretare. Astfel:

a. Prin expresia "scoruri mari (mici)" nu este precizată destul de clar valoarea de scor începând de la care (până la care) sunt valabile adjectivele listei; unii o consideră egală cu m , alții cu $m+\sigma$ ($m-\sigma$), unde m și σ reprezintă media și abaterea standard a scorurilor scalei.

b. Nu este posibilă o diferențiere între subiecții care au obținut scoruri apropiate ca valoare, dacă acestea sunt considerate "mari" sau "mici".

c. Nu există asemenea liste pentru interpretarea scorurilor medii ale scalelor.

Pornind de la aceste considerații am proiectat algoritmi (Albu, 1992a și 1994) care rezolvă ambele probleme formulate anterior, utilizând "funcțiile caracteristice ale adjectivelor" referitoare la o scală sau la o pereche de scale.

13.5.2. O metodă de interpretare a scorurilor în cazul unei singure scale

Se consideră o scală a unui chestionar de personalitate, s , ale cărei scoruri pot lua valorile $0, 1, \dots, n$, și un adjectiv (sau o expresie descriptivă) a .

DEFINIȚIA 13.5.2.1. (Albu, 1992). Prin funcție caracteristică a adjectivului a relativă la scala s se înțelege o funcție

$$f_a : \{0, 1, \dots, n\} \rightarrow [0, 1]$$

care ia ca valoare, în orice punct $x \in \{0, 1, \dots, n\}$, probabilitatea ca adjectivul să fie potrivit pentru caracterizarea unei persoane care a obținut la scala s scorul x .

Algoritmul de construire a funcției caracteristice a adjectivului a relativă la scala s se compune din doi pași:

Pasul 1:

Se obțin punctele experimentale. Pentru aceasta se procedează astfel:

A. Se administrează scala s unui eșantion de subiecți Z , cât mai numeros și mai eterogen, astfel încât să se obțină cât mai multe valori de scor din mulțimea $\{0, 1, \dots, n\}$.

B. Acelorași persoane li se cere să aprecieze dacă adjectivul a li se potrivește sau nu.

C. Se notează:

N = numărul de persoane din lotul experimental, Z ;

S = mulțimea scorurilor $x \in \{0, 1, 2, \dots, n\}$ care au fost obținute la scala s de persoanele din lotul Z .

D. Se definește funcția

astfel:

$$f: S \rightarrow [0,1]$$

$$f(x) = \frac{m(x)}{n(x)}, \quad x \in S \quad (13.5.2.1)$$

unde:

$n(x)$ = numărul de subiecți care au obținut la scala s scorul x :

$$\sum_{x \in S} n(x) = N$$

$m(x)$ = numărul de persoane, dintre cele care au obținut scorul x la scală, care au considerat că adjectivul a li se potrivește;

E. Se reprezintă grafic punctele experimentale $(x, f(x))$, $x \in S$.

Pasul 2:

Se construiește funcția caracteristică a adjectivului a relativă la scala s :

$$f_c: \{0,1,\dots,n\} \rightarrow [0,1]$$

al cărei grafic aproximează punctele experimentale.

Se procedează astfel:

A. Se determină, din mulțimea polinoamelor de grad cel mult trei, funcția de cea mai bună aproximare, în sensul celor mai mici pătrate, a funcției f pe mulțimea S , notată cu P ;

$$P \in \{p: \mathbb{R} \rightarrow \mathbb{R} \mid p(x) = ax^3 + bx^2 + cx + d, a, b, c, d \in \mathbb{R}\}$$

B. Se definește funcția

$$f_c: \{0,1,\dots,n\} \rightarrow [0,1]$$

astfel:

$$f_c(x) = \begin{cases} P(x), & \text{dacă } 0 \leq P(x) \leq 1 \\ 1, & \text{dacă } P(x) > 1 \\ 0, & \text{dacă } P(x) < 0 \end{cases} \quad (13.5.2.2)$$

pentru $x \in \{0,1,\dots,n\}$.

C. Se calculează eroarea medie de reprezentare:

$$E = \frac{1}{N} \sum_{x \in S} n(x) |f(x) - f_c(x)| \quad (13.5.2.3)$$

DEFINIȚIA 13.5.2.2. Se consideră că adjectivul a este adecvat scalei s , la pragul de probabilitate $P_0 \in (0,1)$ fixat și la pragul de eroare fixat ϵ dacă sunt îndeplinite următoarele condiții:

- eroarea medie de reprezentare E este mai mică decât ϵ ;
- funcția caracteristică, f_c , îndeplinește una dintre următoarele condiții:

(C1) f_c este nedescrescătoare, adică:

$$t_1 < t_2 \Rightarrow f_c(t_1) \leq f_c(t_2) \quad (13.5.2.4)$$

(C2) f_c este necrescătoare, adică:

$$t_1 < t_2 \Rightarrow f_c(t_1) \geq f_c(t_2) \quad (13.5.2.5)$$

(C3) există un punct $t_0 \in \{1,2,\dots,n-1\}$ astfel încât, oricare ar fi punctele $t_i \in \{0,1,\dots,n\}$, $i=1,2,3,4$, care îndeplinesc condiția, $t_1 \leq t_0 \leq t_2$ și $t_3 \leq t_0 \leq t_4$, are loc una dintre următoarele situații:

$$f_c(t_1) \leq f_c(t_2) \leq f_c(t_0) \text{ și } f_c(t_0) \geq f_c(t_3) \geq f_c(t_4) \quad (13.5.2.6)$$

sau

$$f_c(t_1) \geq f_c(t_2) \geq f_c(t_0) \text{ și } f_c(t_0) \leq f_c(t_3) \leq f_c(t_4) \quad (13.5.2.7)$$

- funcția f_c are pe mulțimea $\{0,1,\dots,n\}$ valori atât mai mari, cât și mai mici decât P_0 .

Toate adjectivale care sunt adecvate scalei s se rețin pentru a fi utilizate la caracterizarea persoanelor.

Tehnica de selecție a adjectivelor pentru editarea raportului psihologic

Se consideră un chestionar de personalitate ale cărui scale sunt s_1, s_2, \dots, s_m .

Se fixează un prag de probabilitate $P_0 \in (0,1)$ și un prag de eroare ϵ .

Pentru fiecare scală s_i , $i=1,2,\dots,m$, se notează cu A_i mulțimea adjectivelor adecvate scalei s_i la pragul de probabilitate P_0 și la pragul de eroare ϵ .

Pentru o persoană care la scalele chestionarului a obținut scorurile x_1, x_2, \dots, x_m (x_i = scorul la scala s_i , $i=1,2,\dots,m$) editarea raportului psihologic se poate face astfel:

- la fiecare scală s_i , $i=1,2,\dots,m$, se determină valoarea funcției caracteristice a fiecărui adjectiv a din mulțimea A_i , în punctul x_i , $f_{c,i}(x_i)$;
- se rețin adjectivale pentru care $f_{c,i}(x_i) \geq P_0$.

13.5.3. O metodă de interpretare a scorurilor în cazul unei perechi de scale

Se consideră un adjectiv a și două scale ale unui chestionar de personalitate, s_1 și s_2 .

Se notează cu X și Y mulțimile scorurilor posibile la cele două scale.

DEFINIȚIA 13.5.3.1. (Albu, 1994). Prin funcție caracteristică a adjectivului a relativă la perechea de scale (s_1, s_2) înțelegem o funcție

$$F_a: X \times Y \rightarrow [0, 1]$$

care ia ca valoare, în orice punct $(x, y) \in X \times Y$, probabilitatea ca adjectivul a să fie potrivit pentru caracterizarea persoanelor care au obținut scorul x la scala s_1 și scorul y la scala s_2 .

Se notează:

N = numărul de persoane în lotul experimental;

$S \subseteq X \times Y$ = mulțimea perechilor de scoruri (x, y) obținute la cele două scale de persoanele din lotul experimental;

$n(x, y)$ = numărul de subiecți care au obținut scorul x la scala s_1 și scorul y la scala s_2 , $(x, y) \in S$:

$$\sum_{(x, y) \in S} n(x, y) = N$$

$m(x, y)$ = numărul de subiecți, dintre cei care au obținut scorul x la scala s_1 și scorul y la scala s_2 , care au apreciat că adjectivul a li se potrivește, $(x, y) \in S$.

Se definește funcția:

$$F: S \rightarrow [0, 1]$$

astfel:

$$F(x, y) = \frac{m(x, y)}{n(x, y)}, \quad (x, y) \in S \quad (13.5.3.1)$$

Se determină funcția de cea mai bună aproximare, în sensul celor mai mici pătrate, a funcției F , pe mulțimea S , notată cu P :

$$P \in \{p: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R} \mid p(x, y) = ax^3 + bx^2 + cx + dy^3 + ey^2 + fy + g, a, b, c, d, e, f, g \in \mathbb{R}\}$$

Se construiește funcția caracteristică a adjectivului a :

$$F_a: X \times Y \rightarrow [0, 1]$$

astfel:

$$F_a(x, y) = \begin{cases} P(x, y), & \text{dacă } 0 \leq P(x, y) \leq 1 \\ 1, & \text{dacă } P(x, y) > 1 \\ 0, & \text{dacă } P(x, y) < 0 \end{cases} \quad (13.5.3.2)$$

pentru $(x, y) \in X \times Y$

și se calculează eroarea medie de reprezentare, prin formula:

$$E = \frac{1}{N} \sum_{(x, y) \in S} n(x, y) |F(x, y) - F_a(x, y)| \quad (13.5.3.3)$$

DEFINIȚIA 13.5.3.2. Se consideră că adjectivul a este adecvat perechii de scale (s_1, s_2) , la pragul de probabilitate $P_0 \in (0, 1)$ și la pragul de eroare fixat ϵ , dacă funcția sa caracteristică F_a îndeplinește următoarele condiții:

- eroarea medie de reprezentare E este mai mică decât ϵ ;
- restricția funcției F_a pe orice submulțime a mulțimii $X \times Y$ formată din puncte situate pe aceeași dreaptă este sau necrescătoare sau nedescrescătoare sau neconcavă sau neconvexă;
- funcția F_a are pe mulțimea $X \times Y$ valori atât mai mari cât și mai mici decât P_0 .

Toate adjectivele adecvate unei perechi de scale vor fi reținute și vor fi utilizate la descrierea persoanelor.

Pentru caracterizarea unui individ care a obținut la o pereche de scale scorurile (x, y) , vor fi selectate acele adjective ale căror funcții caracteristice satisfac inegalitatea $F_a(x, y) \geq P_0$.

13.5.4. Avantajele celor două metode de interpretare a scorurilor

a. Metodele prezentate permit să se utilizeze, la caracterizarea persoanelor, nu numai adjective corespunzătoare valorilor extreme ale scorurilor scalelor, ci și adjective care sunt potrivite doar scorurilor medii. De exemplu, pentru scala de flexibilitate F_x din CPI adjectivul "adaptabil" este potrivit numai scorurilor medii, la scoruri mari fiind mai potrivit adjectivul "instabil".

b. Ele fac posibilă caracterizarea prin adjective diferite a unor persoane care au scoruri apropiate ca valoare la una sau la două scale.

c. Pot fi utilizate pentru editarea rapoartelor psihologice cu ajutorul calculatorului.

13.5.5. Exemple

Un lot de 436 femei a răspuns la itemii din CPI.
S-au ales pragurile $P_0=0.85$ și $\epsilon=0.20$.

A. S-au construit funcțiile caracteristice ale propoziției:

(p1) "Nu are încredere în forțele proprii"

relative la fiecare dintre scalele Do (dominanță) și In (independentă) din CPI, întrucât această propoziție este inclusă în listele atașate celor două scale, pentru caracterizarea persoanelor care obțin "scoruri mici".

Valorile statistice ale scorurilor celor două scale sunt :

- pentru Do: $m=18.63$ $\sigma=5.76$
- pentru In: $m=15.37$ $\sigma=3.74$

În figurile 13.5.5.1 și 13.5.5.2 sunt reprezentate graficele restricțiilor funcțiilor caracteristice ale propoziției p1 relative la aceste scale, pe mulțimea scorurilor obținute de subiecți.

S-a putut constata că scorul maxim pentru care propoziția p1 este potrivită cu o probabilitate mai mare decât P_0 este mai mic, în ambele cazuri, decât $m-\sigma$: pentru Do acest scor este egal cu 11, iar pentru In, este egal cu 10.

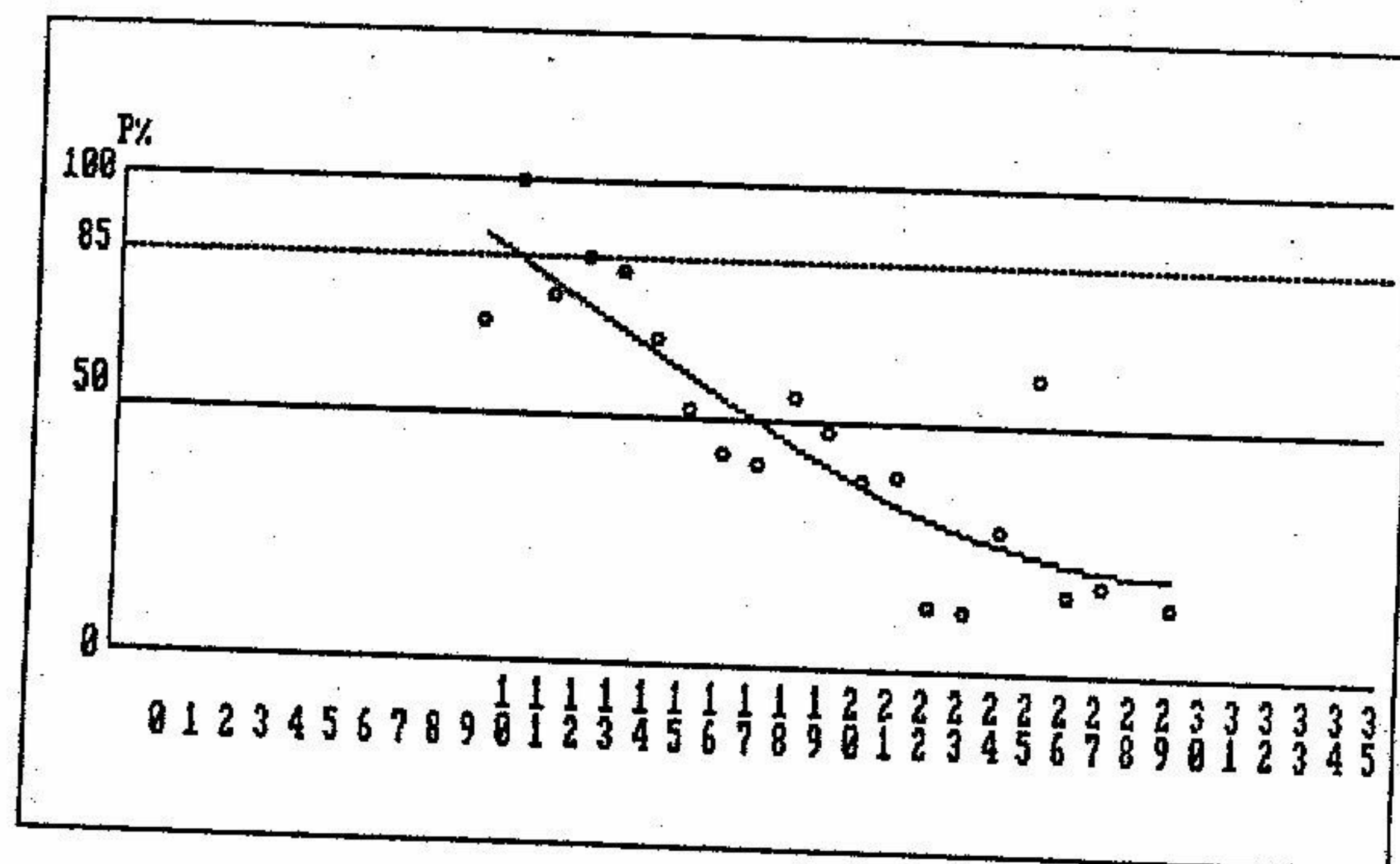


Figura 13.5.5.1

Graficul funcției caracteristice a propoziției p1, relativă la scala Do

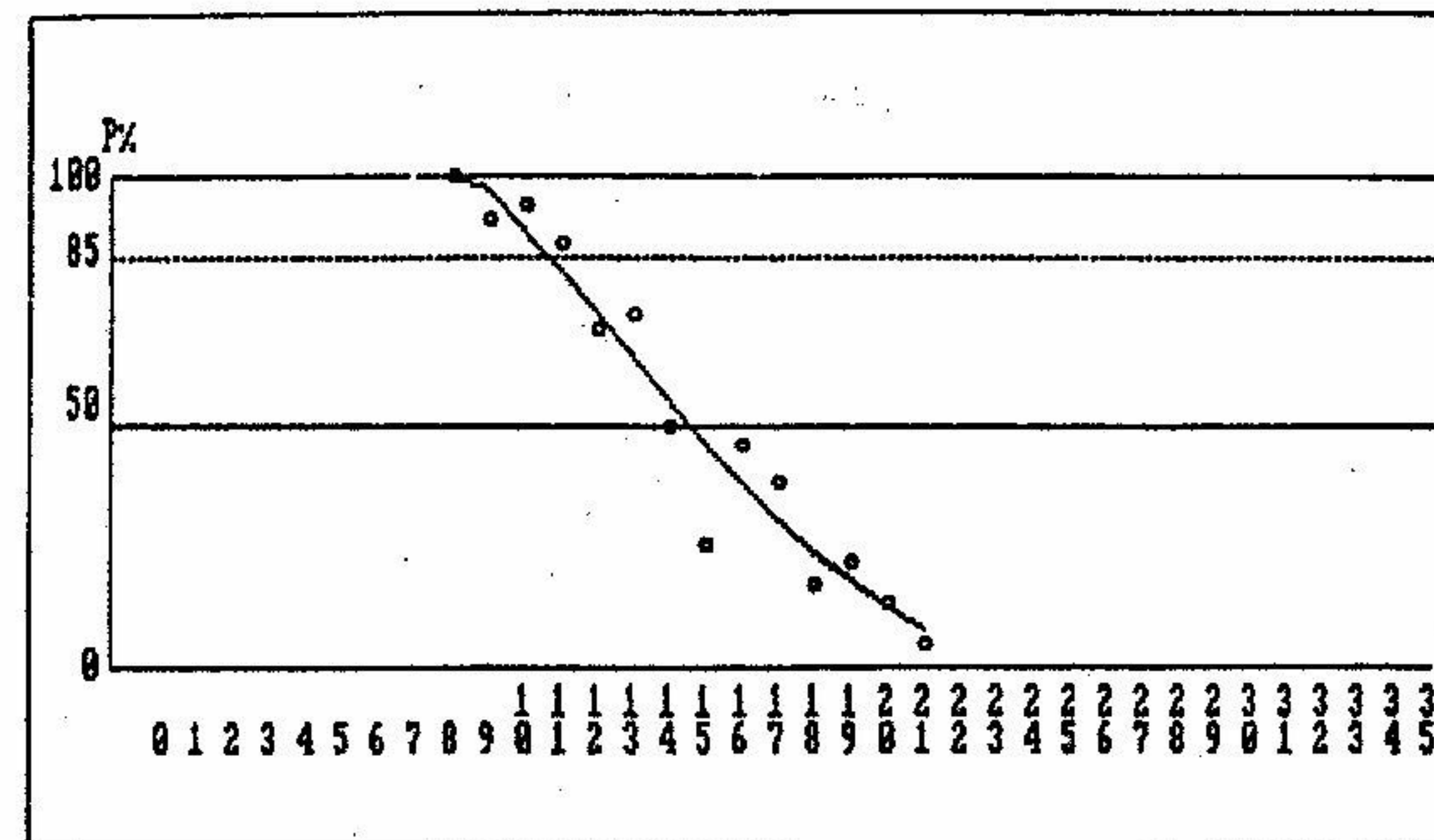


Figura 13.5.5.2

Graficul funcției caracteristice a propoziției p1 relativă la scala In

B. S-au construit funcțiile F_{ϵ} pentru propozițiile:

(p2) Simt o dorință puternică de reușită în viață.

(p3) Mă entuziasmez foarte ușor.

relative la scalele v1 (introversiune / extraversiune) și v2 (acceptarea normelor) din CPI.

Mulțimea S , a perechilor de scoruri obținute la scalele v1 și v2 de lotul experimental, utilizată pentru construirea funcțiilor F_{ϵ} , este inclusă în mulțimea $[12,30] \times [21,33]$. Această mulțime a jucat, în exemplul de față, rolul mulțimii $X \times Y$.

Valorile statistice ale scorurilor celor două scale sunt:

- pentru v1: $m_1=21.17$ $\sigma_1=5.92$
- pentru v2: $m_2=26.84$ $\sigma_2=4.73$

Graficele funcțiilor caracteristice F_{ϵ} pe mulțimea $[12,30] \times [21,33]$ sunt prezentate în figurile 13.5.5.3 (pentru propoziția p2) și 13.5.5.5 (pentru propoziția p3).

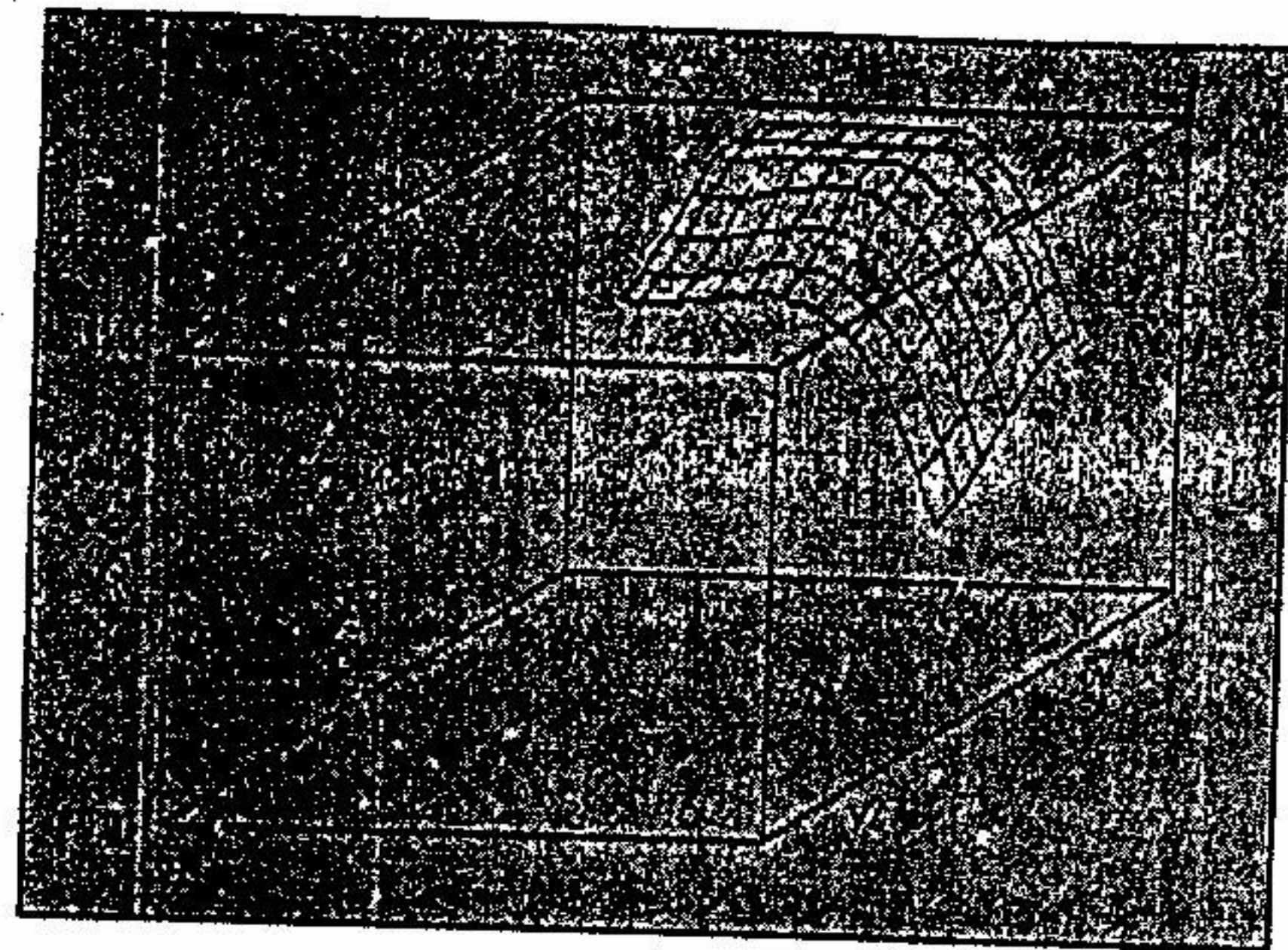


Figura 13.5.5.3
Graficul funcției caracteristice a propoziției $p2$ relativă la perechea de scale $(v1, v2)$

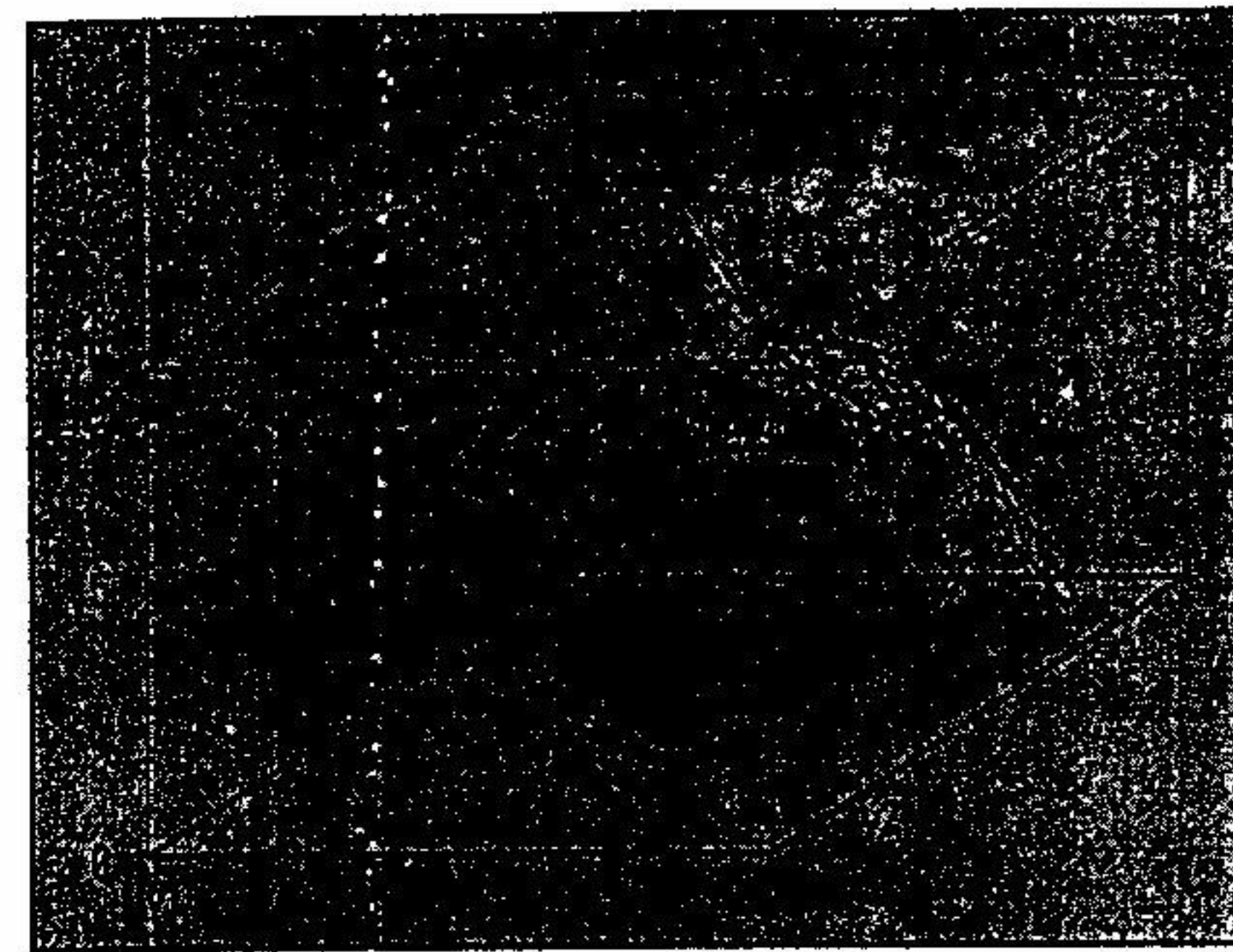


Figura 13.5.5.5
Graficul funcției caracteristice a propoziției $p3$ relativă la perechea de scale $(v1, v2)$

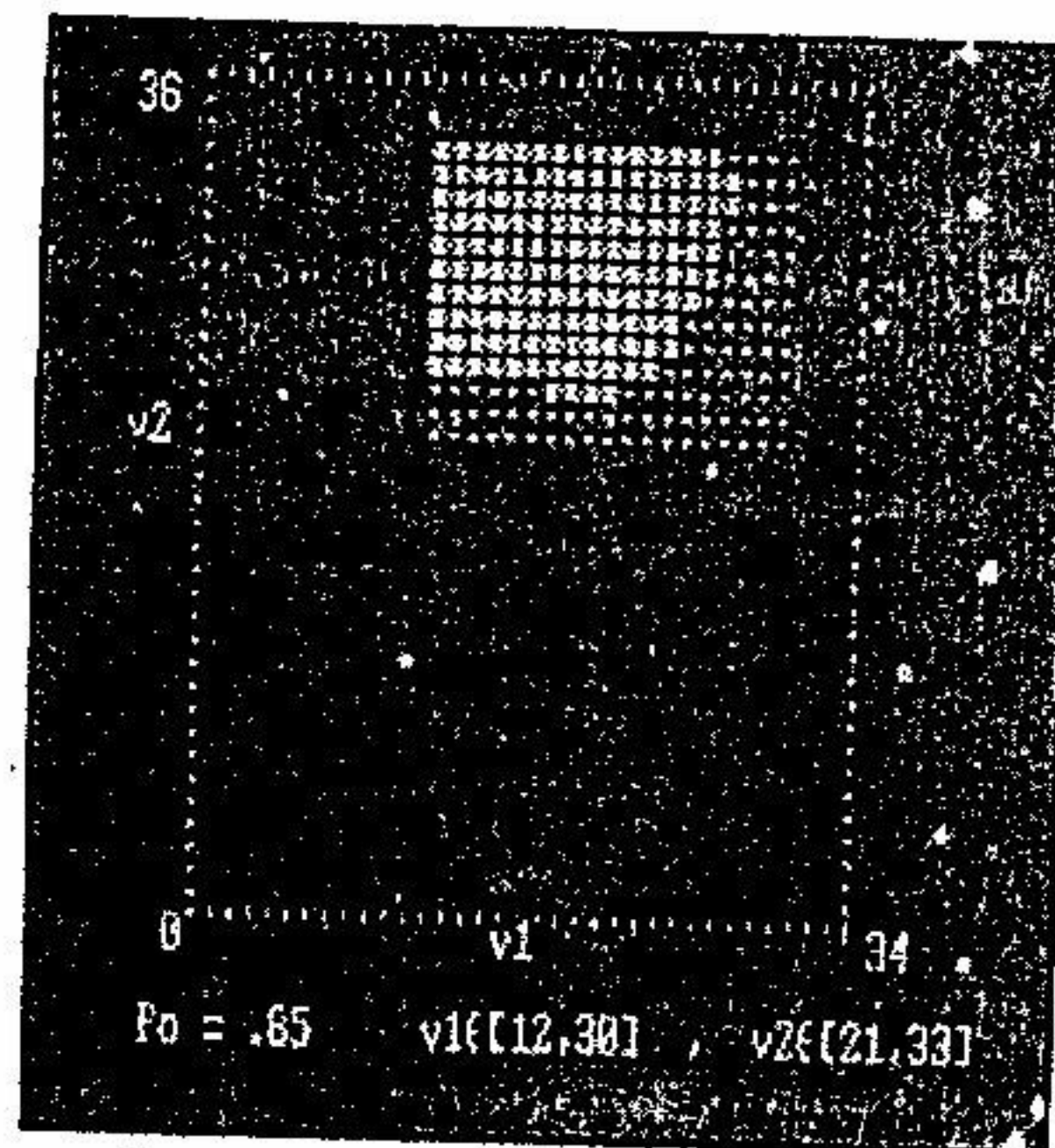


Figura 13.5.5.4
Perechile de scoruri (s_1, s_2) pentru care propoziția $p2$ se poate utiliza în descrierea persoanei

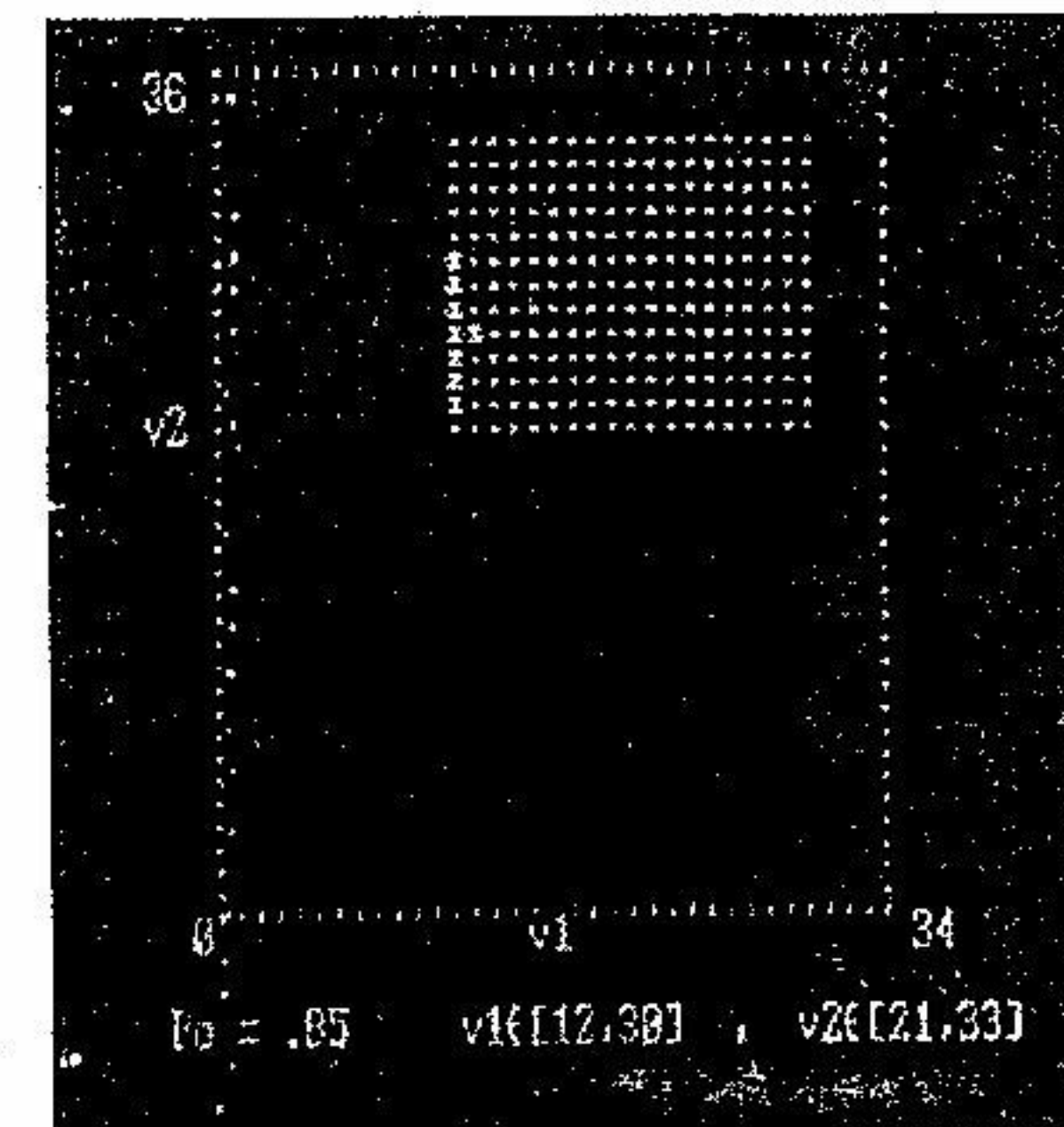


Figura 13.5.5.6
Perechile de scoruri (s_1, s_2) pentru care propoziția $p3$ se poate utiliza în descrierea persoanei

În figurile 13.5.5.4 și 13.5.5.6 sunt notate prin cercuri pline (discuri) perechile de scoruri (s_1, s_2) pentru care valorile funcției caracteristice F_c sunt mai mari decât P_0 .

Considerăm mulțimile formate din perechi de scoruri (s_1, s_2) (s_1 = scorul la scala v1, s_2 = scorul la scala v2), cărora, conform celor arătate în paragraful 13.5.1.1, li se atașează listele 1-4:

$$A1 = \{(s_1, s_2) : s_1 < m_1 - \sigma_1 \text{ și } s_2 \geq m_2 + \sigma_2\}$$

$$A2 = \{(s_1, s_2) : s_1 \geq m_1 + \sigma_1 \text{ și } s_2 \geq m_2 + \sigma_2\}$$

$$A3 = \{(s_1, s_2) : s_1 < m_1 - \sigma_1 \text{ și } s_2 < m_2 - \sigma_2\}$$

$$A4 = \{(s_1, s_2) : s_1 \geq m_1 + \sigma_1 \text{ și } s_2 < m_2 - \sigma_2\}$$

Se observă că, pentru nici una dintre propozițiile $p2$ și $p3$, suprafața pe care funcția caracteristică la valori mai mari decât P_0 nu coincide cu vreuna dintre mulțimile $A1 \cap S$, $A2 \cap S$, $A3 \cap S$, $A4 \cap S$. Aceasta înseamnă că cele două propoziții nu ar trebui incluse în listele atașate perechii de scale (v1, v2). Dar, ele sunt utile pentru caracterizarea persoanelor care au obținut la cele două scale scoruri din mulțimile de puncte notate prin cercuri pline în figurile 13.5.5.4 și 13.5.5.6.

CAPITOLUL 14 PROFILUL PSIHOLAGIC

Frecvent, atât în cercetările științifice cât și în practică, se înregistrează valorile pe care le iau două sau mai multe variabile dependente D_1, D_2, \dots, D_n ($n \geq 2$) pentru diverse valori ale variabilelor independente I_1, I_2, \dots, I_m ($m \geq 1$). Se ajunge la următoarea problemă: dacă valorile variabilelor independente sunt alese astfel încât între oricare două înregistrări cel puțin o variabilă independentă își schimbă valoarea, ce se poate spune despre felul în care se modifică, de la o înregistrare la alta, ansamblul valorilor variabilelor dependente?

În acest capitol sunt discutate câteva posibilități de rezolvare a acestei probleme atunci când variabilele dependente D_1, D_2, \dots, D_n îndeplinesc următoarele condiții:

- sunt măsurate pe scală de același tip (de exemplu, pe scală de interval);
- iau valori în aceeași mulțime (de exemplu, pot lua doar valorile 1, 2, 3, 4 și 5);
- fiecare valoare posibilă are aceeași semnificație pentru toate variabilele (de exemplu, dacă variabilele dependente indică prezența unor atribute, atunci valoarea 0 va însemna "atribut absent", iar valoarea 1, "atribut prezent").

În asemenea situații se obișnuiește să se vizualizeze valorile variabilelor dependente construind câte un profil pentru fiecare înregistrare (Tabachnick, Fidell, 1989).

Trasarea unui profil constă în reprezentarea grafică a punctelor corespunzătoare valorilor variabilelor dependente și unirea punctelor consecutive prin segmente de dreaptă.

Exemple de situații în care se pot realiza profiluri

A. Într-o cercetare referitoare la folosirea calculatoarelor într-o companie, în cursul unei luni, variabilele independente sunt compartimentul companiei și tipul calculatorului, iar variabilele dependente reprezintă timpul mediu de utilizare și timpul mediu de depanare al calculatoarelor. Se poate construi câte un profil pentru fiecare pereche (compartiment, tip de calculator).

B. Un director de instituție își apreciază angajații în privința pregătirii profesionale, a conștiinciozității, a punctualității și a disciplinei. Calificativele acordate

sunt numere întregi, cuprinse între 1 și 5, cu următoarele semnificații: 1=foarte slab, 2=slab, 3=mediocru, 4=bun, 5=foarte bun. În acest caz, variabila independentă corespunde persoanelor din instituție, iar variabilele dependente sunt cele patru criterii utilizate la formularea caracterizărilor. Se pot trasa atâtea profiluri, câte persoane au fost apreciate.

C. Pentru a studia eficacitatea unui medicament se fac măsurători ale unui anumit parametru fiziologic, în cinci momente (înaintea începerii tratamentului și apoi, de patru ori, din 10 în 10 zile), utilizându-se persoane aflate în diferite stadii ale bolii. Variabila independentă are ca valori stadiile bolii, iar variabilele dependente corespund celor cinci momente de măsurare. Fiecare profil va corespunde unui stadiu al bolii. El va conține mediile valorilor parametrului fiziologic, în cele cinci momente, calculate pe baza valorilor notate pentru toate persoanele aflate în stadiul de boală respectiv.

În practică se întâlnesc mai frecvent două moduri de trasare a profilurilor (Piéron, Pichot ș.a., 1952):

- Se folosește un sistem de axe ortogonale, pe una fiind notate variabilele dependente, iar pe cealaltă valorile posibile ale acestora.

- Se desenează câte o rază pentru fiecare variabilă dependentă. Razele au aceeași lungime, pornesc dintr-un centru comun și oricare două raze alăturate formează între ele un unghi egal cu 360° / numărul de variabile dependente. Pe fiecare rază se marchează printr-un punct valoarea variabilei dependente.

Ordinea în care sunt notate într-un profil variabilele dependente este determinată de semnificația acestora. Astfel, dacă variabilele corespund unei caracteristici măsurate sau observate în mai multe momente, ordinea lor va fi cea cronologică. Dacă variabilele dependente pot fi grupate (de exemplu, unele se referă la forma și mărimea unui produs, altele la costurile de producție ale acestuia), atunci cele care fac parte din aceeași grupă vor fi reprezentate alături.

Noțiunea de profil apare și în psihologie, mai ales în orientarea școlară și profesională. Variabilele independente reprezintă persoane sau grupe de subiecți constituite după diverse criterii (de exemplu, sex, vârstă, profesie, nivel de școlarizare etc.). Variabilele dependente sunt rezultatele unor măsurători (de exemplu, timpi de reacție ori frecvențele relative ale unor tipuri de erori), scorurile unor scale sau diferențele de scor între două administrări ale unui test.

Pentru a putea construi profilul psihologic pe baza rezultatelor unor teste se obișnuiește să se transforme cotele brute obținute de subiecți la diverse scale în cote standardizate, în decile sau stanine.

Atunci când profilul corespunde unui grup de persoane, valoarea fiecărei variabile dependente este media valorilor variabilei respective pentru persoanele din grup.

Trasarea profilurilor psihologice servește în principal următoarelor scopuri:

- A. compararea între ele a unor grupe de persoane;
- B. compararea unei persoane cu mai multe grupe, pentru a determina în care grupă ar trebui repartizată;
- C. compararea a două persoane, una cu alta;
- D. compararea unei persoane cu ea însăși, înainte și după un anumit eveniment (de exemplu, un tratament).

În continuare, vom prezenta, pe scurt, procedeele de lucru pentru fiecare dintre aceste scopuri.

A. Compararea profilurilor grupelor utilizează metode statistice specifice, care fac obiectul "analizei de profil". Se încearcă să se dea răspuns următoarelor întrebări:

- a. Cum se asociază variabilele dependente pentru fiecare grup? Ce deosebiri există între grupe în această privință?

- b. Există vreo variabilă dependentă ale cărei valori medii diferă semnificativ între grupe?

- c. Profilurile grupelor sunt paralele între ele?

Consecința paralelismului profilurilor este faptul că scorurile cele mai mici și scorurile cele mai mari apar la toate grupele pentru aceleași variabile. Dacă grupele sunt formate din subiecți supuși unor tratamente diferite și profilurile sunt paralele, se poate deduce că tratamentele au efecte similare.

- d. În cazul profilurilor paralele, există diferențe de "nivel" între grupe? "Nivelul" unui profil este media valorilor tuturor variabilelor dependente.

B. Determinarea grupei din care face parte o persoană este o problemă de clasare, care se rezolvă adesea prin compararea persoanei respective cu prototipurile claselor și identificarea prototipului cu care aceasta "seamănă" cel mai mult.

Atunci când profilurile grupelor sunt diferite, se poate considera fiecare profil ca fiind prototip al grupei. Se determină profilul persoanei, pe baza valorilor variabilelor dependente care au fost utilizate la trasarea profilurilor grupelor, și se compară profilul persoanei cu profilul fiecărei grupe.

Exemplu.

Pentru a afla dacă o persoană are tulburări socio-emoționale se pot utiliza drept variabile dependente scalele chestionarelor STAI și STAXI. Se compară scorurile obținute de subiect cu scorurile medii realizate de două loturi: unul format din persoane cu tulburări socio-emoționale și unul compus din persoane fără asemenea tulburări.

De obicei, compararea profilului persoanei cu profilul unei grupe constă în calculul distanței dintre punctele

$$(p_1, p_2, \dots, p_n) \text{ și } (g_1, g_2, \dots, g_n),$$

unde, pentru $j=1, 2, \dots, n$:

p_j este valoarea variabilei dependente D_j la persoana respectivă;

g_j este media valorilor variabilei dependente D_j în cadrul grupei.

Pentru calculul distanței dintre punctele (p_1, p_2, \dots, p_n) și (g_1, g_2, \dots, g_n) se va utiliza una dintre distanțele spațiului R^n (de exemplu, distanța euclidiană, distanța lui Mahalanobis, distanța city-block etc.), aleasă în funcție de importanța acordată distanțelor dintre valorile componentelor (a se vedea Anexa II).

Asemănarea profilurilor este cu atât mai mare, cu cât distanța dintre puncte este mai mică.

Compararea unui profil individual cu profilul unui grup impune să se aibă în vedere și următoarele aspecte:

a. Rezultatele comparării profilurilor depind de funcția aleasă pentru distanță.

b. O aceeași valoare a distanței poate să apară atunci când între profilurile comparate diferă mult valorile unei singure variabile dependente, celelalte variabile având valori egale, sau când toate variabilele dependente diferă puțin între profiluri. Sunt importante și diferențele mici între valorile variabilelor dependente sau acestea ar trebui neglijate? Are sau nu importanță care este variabila dependentă pentru care apar diferențe mari între valori?

c. În funcțiile utilizate pentru calculul distanțelor din spațiul R^n nu au importanță semnele diferențelor $p_j - g_j$ ($j=1, 2, \dots, n$) ci doar mărimile lor în valoare absolută. Dar, în problemele practice, nu contează și semnele diferențelor? De exemplu, dacă se face compararea profilurilor pentru a identifica profesia cea mai potrivită unei persoane, are importanță dacă în privința unei variabile individul respectiv este mai bun sau mai slab decât media grupei ori contează doar mărimea diferenței dintre rezultatul său și media grupei, în valoare absolută?

C și D. Pentru a putea interpreta corect diferențele dintre două profiluri individuale, care aparțin la doi subiecți diferiți sau aceluiași subiect în două momente diferite, ar trebui să se poată ști dacă diferența valorilor observate la fiecare variabilă dependentă corespunde unei deosebiri reale sau este datorată întâmplării.

Prezentăm o metodă de comparare a două profiluri individuale, care pornește de la ideea că orice scor observat la un test este afectat de erori de măsură și ia în considerare, pentru fiecare variabilă dependentă, scorurile care ar fi putut să apară, cu o anumită probabilitate fixată, în locul celui observat. Această metodă este aplicabilă atunci când sunt îndeplinite următoarele condiții:

• Variabilele dependente reprezintă scoruri brute la teste care măsoară caracteristici durabile.

• Pentru fiecare test scorurile posibile aparțin unei mulțimi finite formate din numere

întregi consecutive:

$X = \{x_1, x_2, \dots, x_n\}$ unde $x_i \in \mathbb{Z}$ și $x_i = x_{i-1} + 1$, $i=2, 3, \dots, n$ (aceste mulțimi pot să difere de la un test la altul).

• Fiecare test a fost administrat de două ori aceluiași lot de subiecți care, în privința variabilelor dependente studiate nu s-au modificat între test și retest. Verificarea acestei condiții se face urmărind dacă relația dintre scorurile testului și cele ale retestului este liniară, semnificativă statistic la un prag p fixat ($p=0.01$ sau $p=0.001$), și dacă diferența mediilor scorurilor între test și retest este ne semnificativă statistic la pragul p .

Se procedează astfel:

(1) Se fixează o probabilitate $P \in [0, 1]$.

(2) Pentru fiecare test se rezolvă problema următoare (Albu, Pitariu, 1995b): să se determine, pentru fiecare scor posibil al testului, $x \in X$, un interval care să îl conțină:

$$[u(x, P), v(x, P)], \quad u(x, P) \leq x \leq v(x, P),$$

astfel încât să fie îndeplinite următoarele condiții:

(C1) Probabilitatea ca o persoană care a obținut la o administrare a testului x puncte să realizeze la o altă administrare (anterioară sau posterioară celei la care subiectul a fost cotat cu x puncte) y puncte, $y \in [u(x, P), v(x, P)]$, să fie cel puțin egală cu P :

$$P(y \in [u(x, P), v(x, P)] | x) \geq P \quad (14.1)$$

Intervalul $[u(x, P), v(x, P)]$ conține scorurile "cele mai apropiate probabilistic" de x .

(C2) Pentru orice scor posibil $x \in X$, dacă $y \in [u(x, P), v(x, P)]$, atunci și $x \in [u(y, P), v(y, P)]$ (dacă y este unul dintre scorurile "cele mai apropiate probabilistic" de x , atunci și x este unul dintre scorurile "cele mai apropiate probabilistic" de y);

(C3) Extremitățile intervalelor $[u(x, P), v(x, P)]$ să fie cât mai apropiate de x , pentru toate scorurile posibile x :

$$C(P) = \max_x \max \{x - u(x, P), v(x, P) - x\} \rightarrow \min \quad (14.2)$$

Trebuie remarcat că problema determinării intervalelor $[u(x, P), v(x, P)]$, pentru $x \in X$, care să respecte condițiile (C1) și (C2), are întotdeauna soluție, căci

se poate construi, pentru fiecare scor posibil x , intervalul $[u(x,P), v(x,P)]$ punând $u(x,P)=x_i$ și $v(x,P)=x_n$.

Problema construirii intervalelor $[u(x,P), v(x,P)]$, pentru $x \in X$, care să respecte condițiile (C1), (C2) și (C3), atunci când testul a fost administrat de două ori unui lot de persoane și fiecare scor posibil a fost obținut cel puțin o dată, poate fi transpusă matematic sub forma unei probleme de programare.

Se notează:

f_{ij} = numărul de subiecți care au obținut scorul x_i la test și scorul x_j la retest, $i, j=1, 2, \dots, n$;

Faptul că fiecare scor posibil a fost obținut cel puțin o dată (indiferent la care dintre administrări) se exprimă prin inegalitatea:

$$\sum_{j=1}^n (f_{ij} + f_{ji}) - f_{ii} > 0, \quad i=1, 2, \dots, n \quad (14.3)$$

Introducem variabilele a_{ij} , $i, j=1, 2, \dots, n$, cu următoarea semnificație:
 $a_{ij}=1$ dacă scorul x_i aparține intervalului construit pentru scorul x_j și
 $a_{ij}=0$ în caz contrar.

Se cere să se determine numerele a_{ij} , $i, j=1, 2, \dots, n$, în următoarele condiții:

(a) $a_{ij} \in \{0, 1\}$, $i, j=1, 2, \dots, n$ (din construcția variabilelor a_{ij}) (14.4)

(b) $a_{ii}=1$, $i=1, 2, \dots, n$ (fiecare scor posibil x trebuie să facă parte din intervalul "scorurilor care sunt cele mai apropiate probabilistic" de el) (14.5)

(c) $a_{ij}=a_{ji}$, $i, j=1, 2, \dots, n$ (condiția C2 a problemei) (14.6)

(d) $a_{i,1} \leq a_{i,2} \leq \dots \leq a_{i,i-1} \leq a_{i,i}$ (14.7)

$a_{i,i} \geq a_{i,i+1} \geq \dots \geq a_{i,n-1} \geq a_{i,n}$ (14.8)

pentru $i=1, 2, \dots, n$
 (pentru a avea câte un interval $[u(x,P), v(x,P)]$ în jurul fiecărui scor posibil x)

(e)

$$P(x_i) = \frac{\sum_{j=1}^n a_{ij} f_{ij} + \sum_{j=1}^n a_{ji} f_{ji} - f_{ii}}{\sum_{j=1}^n (f_{ij} + f_{ji}) - f_{ii}} \geq P \dots \quad (14.9)$$

pentru $i=1, 2, \dots, n$.

(condiția C1 a problemei: $P(x_i)$ reprezintă frecvența relativă a subiecților care au obținut la o administrare a testului scorul x_i , iar la cealaltă un scor din intervalul construit pentru x_i , din totalul celor care au obținut la una dintre administrările testului scorul x_i)

(f)

$$C(P) = \max_{1 \leq i \leq n} \max \left\{ \sum_{j=i+1}^n a_{ij}, \sum_{j=1}^{i-1} a_{ij} \right\} \rightarrow \min \quad (14.10)$$

(condiția C3: prima sumă reprezintă diferența $v(x,P)-x_i$, iar a doua, diferența $x_i-u(x,P)$, pentru $x=x_i$)

(3) Pentru fiecare din cele două cazuri în care se trasează profilul, la fiecare test, se înlocuiește scorul obținut, x , cu intervalul $[u(x), v(x)]$, în care se află scorurile care, cu probabilitatea P , ar fi putut să apară în locul scorului x . Se înlocuiește astfel profilul tradițional, în care pentru fiecare test este desenat un punct, corespunzător scorului obținut de subiect, cu un profil în care pentru fiecare test este trasat segmentul corespunzător intervalului în care, cu probabilitatea P , s-ar fi putut găsi scorul subiectului.

(4) La un test la care valorile în cele două profiluri sunt x și y se va afirma că profilurile diferă numai dacă intervalele corespunzătoare celor două scoruri, $[u(x), v(x)]$ și $[u(y), v(y)]$ au intersecția vidă.

Exemplu.

În figura 14.1 sunt prezentate profilurile psihologice la CPI pentru două persoane. Fiecare profil s-a obținut prin reprezentarea grafică a cotelor T la primele 20 de scale din CPI.

Aplicând procedeul descris mai sus, s-a determinat, pentru probabilitatea $P=0.85$, intervalul în care s-ar fi putut afla scorul brut al fiecărei scale. Limitele intervalelor au fost transformate în cote T . În figura 14.2 sunt prezentate profilurile celor două persoane, realizate cu ajutorul intervalelor. Se observă că pentru unele scale (Do, In, Sc, Gi, Wb, Ac) intervalele celor două profiluri au porțiuni comune, deci este posibil ca la scalele respective persoanele să nu se deosebească una de alta.

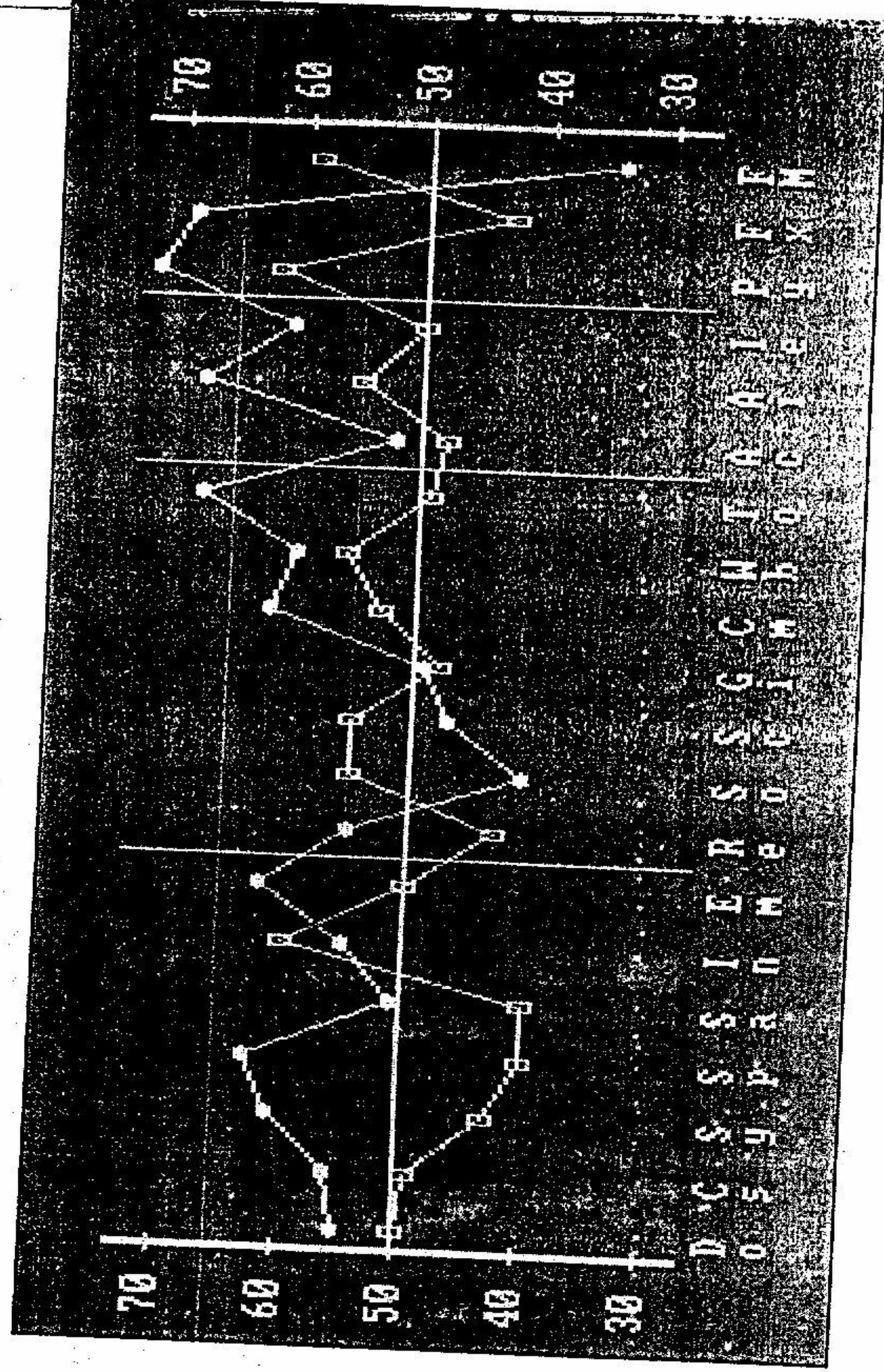


Figura 14.1
 Profilurile psihologice la CPI, reprezentate cu ajutorul cotelor T, pentru două persoane

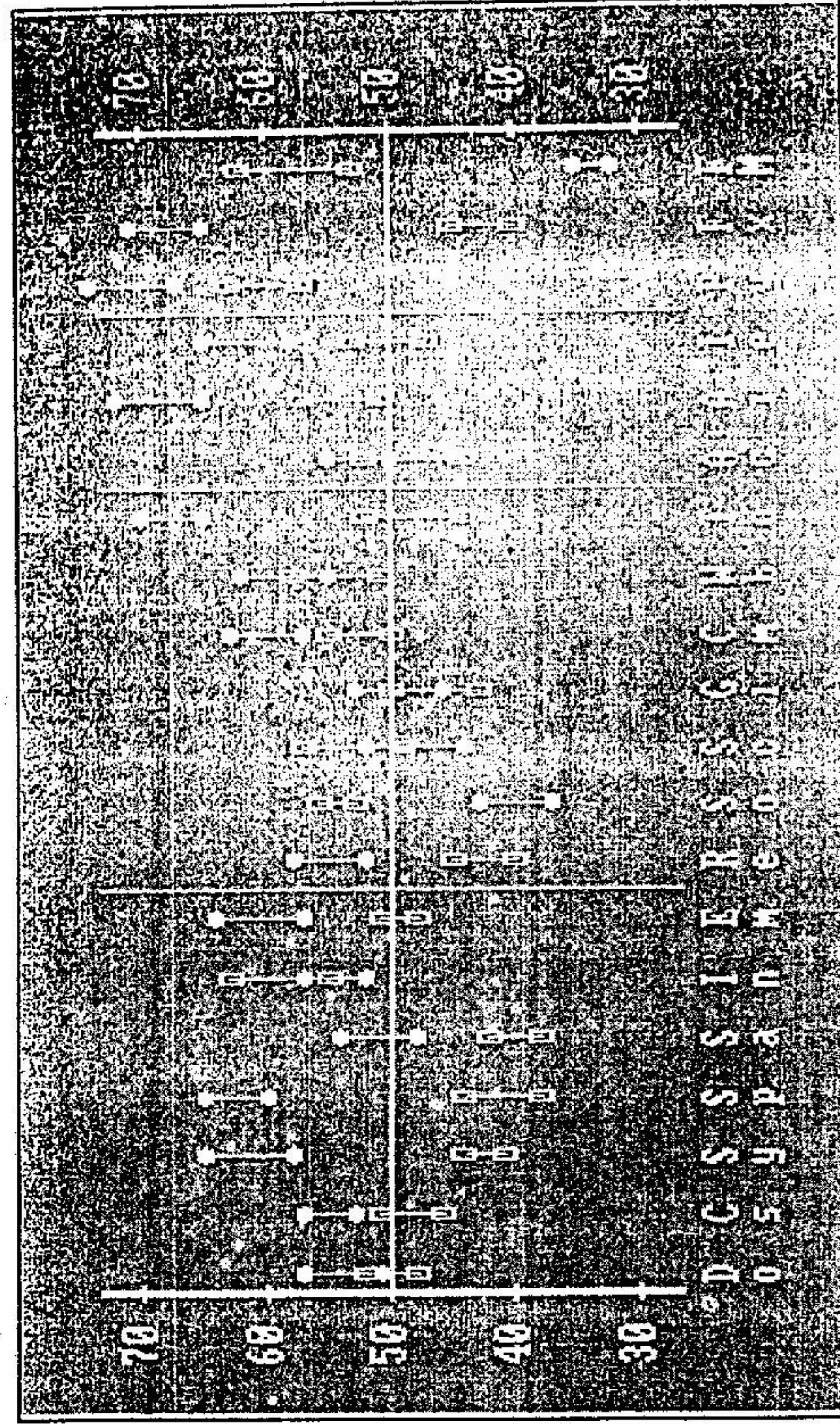


Figura 14.2
 Profilurile psihologice la CPI, reprezentate cu ajutorul intervalelor corespunzătoare probabilității $P=0.85$, pentru două persoane

CAPITOLUL 15

O ABORDARE A TEORIEI TESTELOR PSIHOLOGICE BAZATĂ PE TEORIA CLASIFICĂRII

Înțelegerea acestui capitol necesită parcurgerea prealabilă a Anexei I.

15.1. TESTUL PSIHOLOGIC CA INSTRUMENT DE CLASIFICARE

Utilizarea în practică a testelor psihologice revine la rezolvarea unor probleme de clasificare. De exemplu:

A. Ordönarea persoanelor pe baza scorurilor obținute la un test înseamnă formarea unui *clasament* în mulțimea subiecților.

B. Alegerea, dintr-o mulțime de candidați examinați cu mai multe probe psihologice, a celui mai potrivit pentru un anumit post impune *clasarea* candidaților în una din clasele disjuncte "admis" și "respins", urmată de *ierarhizarea* celor admiși pe baza rezultatelor la probele psihologice.

C. Selecția sau orientarea profesională se realizează conform unui algoritm compus din următorii pași:

(1) Se stabilesc clasele în care se va face clasarea candidaților. O clasă corespunde unui post sau unei grupe de posturi asemănătoare prin cerințe.

(2) Se determină variabilele, fizice sau/și psihice, care permit o diferențiere între posturile pentru care se face selecția. De exemplu, pentru un post poate fi necesară rezistența fizică, iar pentru altul, o atenție distributivă bună.

(3) Se aleg instrumentele care măsoară aceste variabile și permit diferențieri între clase, în sensul că valorile lor diferă de la o clasă la alta.

(4) Pentru fiecare clasă se descrie un *prototip*, folosind valori posibile sau ideale ale măsurătorilor făcute cu instrumentele alese.

(5) Fiecare persoană examinată cu instrumentele alese este comparată cu toate prototipurile. Ea va fi *clasată* în clasa prototipului cu care se aseamănă cel mai mult.

15.2. CONSTRUIREA TESTELOR PSIHOLOGICE

Calculul scorurilor unui test prin însumarea, ponderată sau nu, a scorurilor itemilor componenți, reprezintă un procedeu de agregare a unor clasamente:

- Fiecare item, prin scorurile sale, definește o relație de preordine totală în mulțimea E a subiecților.

- Scorul testului, la rândul său, definește o relație de preordine totală pe aceeași mulțime E și, deci, conduce la un clasament.

Construirea unui test psihologic urmărește să selecteze, dintr-o mulțime dată de itemi, un număr (fixat a priori sau nu) de itemi și să stabilească un procedeu de agregare a clasamentelor (concret, de calcul al scorului testului) astfel încât clasamentul rezultat, corespunzător scorurilor testului, să fie cât mai asemănător cu un anumit clasament existent, cunoscut sau nu. Clasamentul existent este dat de scorurile unui criteriu sau de valorile unei variabile latente. În cel de-al doilea caz, clasamentul nu este cunoscut. Adesea se presupune că el coincide cu clasamentul realizat de scorurile factoriale ale subiecților pentru un factor comun itemilor.

Vom explica în continuare ideile care stau la baza procedeele de construire a testelor.

Se consideră o mulțime de persoane, E , și doi itemi de test. Pentru fiecare persoană care a răspuns la ambii itemi se calculează un "scor total", s_i , folosind formula:

$$s_i = w_1 s_{i1} + w_2 s_{i2} \quad (15.2.1)$$

unde:

s_{i1}, s_{i2} sunt scorurile persoanei la cei doi itemi;

w_1, w_2 sunt ponderi (numere reale pozitive) constante pentru toate persoanele.

Scorul total calculat astfel realizează un clasament în mulțimea E .

Procedeele statistice aplicate pentru construirea testelor psihologice au la bază următoarele teoreme:

TEOREMA 15.2.1. Dacă cele două clasamente corespunzătoare itemilor coincid, atunci clasamentul corespunzător scorului total coincide cu ele.

TEOREMA 15.2.2. Atunci când clasamentul corespunzător primului item este o rafinare a clasamentului corespunzător celui de-al doilea item, clasamentul corespunzător scorului total coincide cu clasamentul corespunzător primului item.

TEOREMA 15.2.3. Dacă cele două clasamente corespunzătoare itemilor coincid cu un clasament dat, atunci ele coincid între ele.

TEOREMA 15.2.4. Atunci când clasamentul corespunzător primului item coincide cu clasamentul corespunzător celui de-al doilea item, care este o rafinare a unui clasament dat, clasamentul corespunzător primului item este și el o rafinare a clasamentului dat.

Aceste teoreme se aplică succesiv, luând în locul celor doi itemi, testul existent la un moment dat și un item din mulțimea celor care pot fi adăugați testului.

În practica psihologică, însă, se întâlnesc foarte rar clasamente care să coincidă sau unul să fie o rafinare a altuia. Intervin diverse erori de măsură care au drept consecință faptul că persoane care obțin același scor la un item (și, deci, aparțin aceleiași clase în partiția ordonată corespunzătoare itemului) obțin scoruri diferite la un alt item (făcând parte din clase diferite în partiția ordonată corespunzătoare acestuia). Din acest motiv nu se mai caută clasamente astfel încât unul să fie o rafinare a altuia sau să coincidă cu el, ci clasamente care să "concorde" între ele. Concret, ce se înțelege prin "concordanță", diferă de la o metodă statistică la alta. Dar, în esență, concordanța a două clasamente se referă la apropierea lor de situația când ele coincid (dacă relațiile de preordine totală cărora le corespund clasamentele au același număr de clase de echivalență) sau când unul este o rafinare a celuilalt (dacă numărul de clase de echivalență ale relațiilor de preordine totală diferă).

În cadrul analizei de itemi, al metodelor de construire a testelor și al cercetării validității relative la criteriu se folosesc diverse formule pentru evaluarea concordanței a două clasamente.

Astfel, studiul validității relative la criteriu a testelor și al validității itemilor compară clasamentul corespunzător unui criteriu cu cel corespunzător testului sau itemului cu ajutorul unui coeficient de corelație (corelație liniară, corelație biserială sau corelația rangurilor).

În cercetarea puterii de discriminare a itemilor, clasificarea corespunzătoare criteriului se reduce la trei clase, din care, pentru verificarea concordanței clasamentelor se folosesc doar cele două extreme ("contrastante"). Itemul are putere de discriminare dacă scorurile sale realizează un clasament concordant cu cel dat de clasele contrastante.

Eliminarea itemilor al căror indice de dificultate este egal sau aproximativ egal cu 0 sau cu 1 are ca explicație faptul că un clasament în care toate persoanele ocupă același loc nu poate să concorde cu nici un clasament în care există mai multe locuri.

Construirea unui test care să aibă o validitate bună relativ la un criteriu, prin selectarea acelor itemi care au o validitate bună față de criteriul respectiv, este de asemenea o aplicație a teoremelor 15.2.1 și 15.2.3, în care termenul "coincid" a fost înlocuit cu termenul "concordă". Astfel: atunci când clasamentele corespunzătoare celor doi itemi concordă cu clasamentul corespunzător criteriului, clasamentele lor concordă între ele, va concorda și cu clasamentul corespunzător scorului total care, la rândul său, va concorda cu clasamentul corespunzător criteriului.

Importanța introducerii în test doar a itemilor valizi rezultă din următorul exemplu.

Exemplu.

Să considerăm 10 persoane, s_1, s_2, \dots, s_{10} , pentru care se cunosc valorile unui criteriu. Să presupunem că scorurile criteriului verifică inegalitățile:

$$c(s_1) < c(s_2) < \dots < c(s_{10}) \quad (15.2.2)$$

Se dorește să se construiască un test care să fie valid față de acest criteriu.

Se pornește de la o colecție inițială formată din 6 itemi dihotomici, i_1, i_2, \dots, i_6 . Scorurile obținute de subiecți la acești itemi și la testul compus din ei sunt notate în Tabelul 15.2.1.

Tabelul 15.2.1
Scorurile itemilor i_1, i_2, \dots, i_6 pentru
subiecții s_1, s_2, \dots, s_{10}

Item	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}
i_1	0	0	0	0	0	0	0	0	1	1
i_2	0	0	0	0	0	0	0	1	1	1
i_3	0	0	0	1	1	1	1	1	1	1
i_4	0	0	1	1	1	1	1	0	0	0
i_5	1	1	1	1	1	1	0	0	0	0
i_6	1	1	1	1	0	0	0	0	0	0
Scor total	2	2	3	4	3	3	2	2	3	3

Analizând scorurile itemilor și ale testului se deduc următoarele:

• Pentru primii trei itemi partițiile ordonate corespunzătoare scorurilor lor sunt rafinate de partiția ordonată corespunzătoare scorurilor criteriului. Dacă un subiect s_i este inferior altui subiect s_j la criteriu, $c(s_i) < c(s_j)$, atunci s_i nu obține la item un scor mai mare decât s_j ($i, j \in \{1, 2, \dots, 10\}$).

• Itemii i_5 și i_6 produc clasamente "inverse": dacă s-ar schimba modul lor de cotare (s-ar da 1 punct în loc de 0 puncte și invers), ar rezulta clasamente pentru care clasamentul bazat pe valorile variabilei criteriu ar fi o rafinare.

• Itemul i_4 este necorespunzător: obțin scorul 1 la item persoanele care au rezultate medii la criteriu și, prin urmare, clasamentul produs de el nu coincide și nu este rafinat de clasamentul produs de criteriu.

• Testul format din cei șase itemi produce o ierarhie foarte diferită de cea bazată pe variabila criteriu. De exemplu, scorul maxim la test îl are subiectul s_4 care nu se află printre cei cu valorile cele mai mari ale criteriului.

Dar, ținând cont de rezultatele analizei validității itemilor, se poate construi un test care să nu modifice prea mult ordinea subiecților. Pentru aceasta se schimbă modul de cotare pentru itemii i_5 și i_6 (scor nou := 1 - scor vechi) și se renunță la itemul i_4 . Scorurile noi ale itemilor și scorul total al testului sunt prezentate în tabelul 15.2.2.

Tabelul 15.2.2
Scorurile noi ale itemilor și scorul testului pentru
subiecții s_1, s_2, \dots, s_{10}

Item	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}
i_1	0	0	0	0	0	0	0	0	1	1
i_2	0	0	0	0	0	0	0	1	1	1
i_3	0	0	0	1	1	1	1	1	1	1
i_5	0	0	0	0	0	0	1	1	1	1
i_6	0	0	0	0	1	1	1	1	1	1
Scor total	0	0	0	1	2	2	3	4	5	5

În teoria răspunsului la itemi se consideră că un item dihotomic este valid dacă are curba caracteristică (curba de regresie) crescătoare față de variabila măsurată de test. Alura curbei de regresie constituie, deci, o metodă de verificare a concordanței între clasamentul corespunzător scorurilor itemului și cel corespunzător valorilor variabilei măsurate de itemi (acest al doilea clasament nu este cunoscut).

Și atunci când testul se construiește utilizând teoria răspunsului la itemi, obținerea unui clasament corect al subiecților este condiționată de alegerea itemilor valizi.

Exemplu (Albu, 1993).

Considerăm un test alcătuit din trei itemi dihotomici, care măsoară aceeași trăsătură psihică și ale căror funcții caracteristice au expresia analitică:

$$P(y) = c + \frac{1-c}{1+e^{-1.7a(y-b)}} \quad (15.2.3)$$

Parametrii funcțiilor caracteristice sunt:

$$\begin{array}{lll} a_1=1 & b_1=6 & c_1=0 \\ a_2=-0.1 & b_2=15 & c_2=0 \\ a_3=1 & b_3=11 & c_3=0 \end{array}$$

Funcția caracteristică a celui de-al doilea item este descrescătoare ($a_2 < 0$), deci itemul nu este valid.

La fiecare nivel al variabilei măsurate, $t \in \mathbb{R}$, sunt posibile, teoretic, toate scorurile testului. Luând în considerare probabilitățile scorurilor, există în fiecare punct $t \in \mathbb{R}$ o repartiție de probabilitate de următoarea formă:

$$\begin{pmatrix} 0 & 1 & 2 & 3 \\ P_0(t) & P_1(t) & P_2(t) & P_3(t) \end{pmatrix} \quad (15.2.4)$$

cu $P_i(t) \geq 0$, $i=0,1,2,3$
și $P_0(t)+P_1(t)+P_2(t)+P_3(t)=1$.

Valorile $P_0(t)$, $P_1(t)$, $P_2(t)$ și $P_3(t)$ se pot calcula cu ajutorul funcțiilor caracteristice ale itemilor.

Notăm cu $A_i(t)$ evenimentul ca itemul i să fie cotat cu un punct pentru o persoană având nivelul variabilei măsurate egal cu t , pentru $i=1,2,3$. Probabilitatea evenimentului $A_i(t)$ este egală cu valoarea în punctul t a funcției caracteristice a itemului, $p_i(t)$.

Acceptăm ipotezele obișnuite din teoria răspunsului la itemi (Lord, 1980):

- variabila măsurată ia valori pe întreaga axă reală;
- în fiecare punct $t \in \mathbb{R}$, evenimentele $A_i(t)$, $i=1,2,3$, sunt independente probabilistic între ele.

În acest caz se pot scrie formule de calcul pentru probabilitățile P_i , $i=0,1,2,3$. Acestea sunt (pentru simplificarea scrierii renunțăm la notarea argumentului t):

$$\begin{aligned}
 P_0 &= (1-p_1)(1-p_2)(1-p_3) \\
 P_1 &= p_1(1-p_2)(1-p_3) + (1-p_1)p_2(1-p_3) + (1-p_1)(1-p_2)p_3 \\
 P_2 &= (1-p_1)p_2p_3 + p_1(1-p_2)p_3 + p_1p_2(1-p_3) \\
 P_3 &= p_1p_2p_3
 \end{aligned}$$

S-a studiat testul format din cei trei itemi cu ajutorul funcției care ia ca valoare, în fiecare punct $t \in \mathbb{R}$, scorul cel mai probabil la test, $S_{\max}(t)$:

$S_{\max} : \mathbb{R} \rightarrow \{0, 1, 2, 3\}$
 $S_{\max}(t) = \max \{j \mid P_j(t) = \max \{P_i(t) \mid i=0, 1, 2, 3\}\}$
 ($S_{\max}(t)$ este scorul de test care are cea mai mare probabilitate în punctul t ; dacă există mai multe valori pentru care probabilitatea este egală cu cea maximă, atunci $S_{\max}(t)$ este cea mai mare dintre ele.)

Graficul funcției S_{\max} este prezentat în figura 15.2.1. După cum se observă, scorul cel mai probabil al testului nu mai este nedescrescător față de caracteristica măsurată de test. Persoanele pentru care scorul cel mai probabil este scorul maxim ($S_{\max}=3$) sunt cele cu un nivel mediu al variabilei măsurate de test.

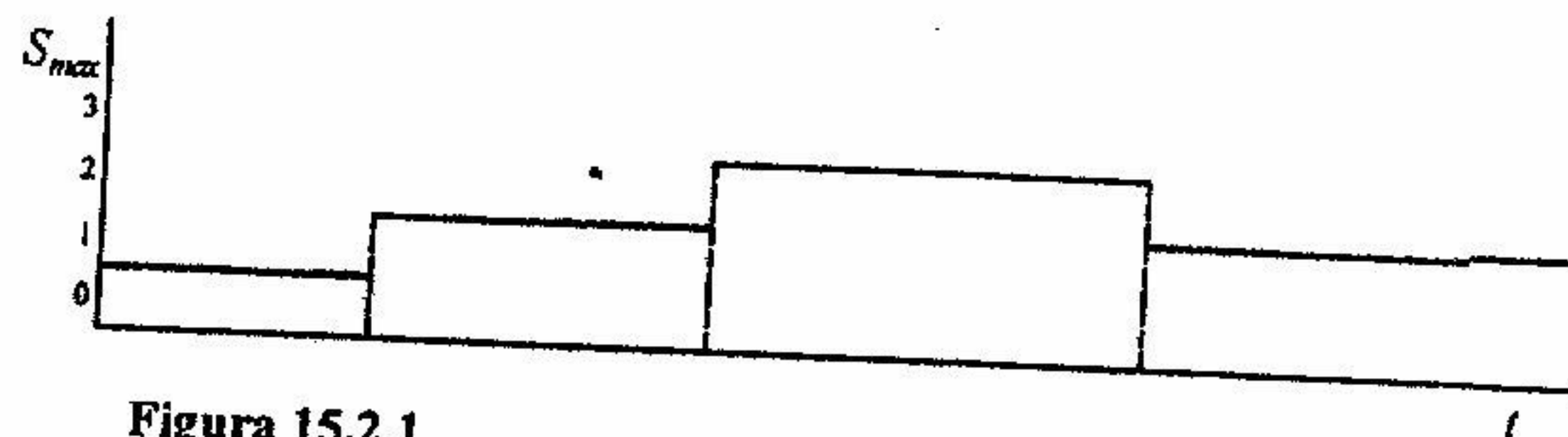


Figura 15.2.1
Graficul funcției S_{\max}

Din exemplele prezentate se deduce că, pentru a obține un test care să realizeze un clasament al subiecților concordant cu cel dat de un criteriu sau de o variabilă latentă este necesar să se analizeze validitatea itemilor și să se rețină în test doar itemi valizi.

Metoda analizei factoriale construiește un test psihologic selectând itemii cu saturații mari într-un același factor, adică itemii care corelează liniar semnificativ cu un același factor. O altă metodă de construire a testelor reține în test doar itemii care corelează liniar semnificativ unii cu alții. Ideea care stă la baza acestor procedee este că există o variabilă latentă ale cărei valori produc un clasament cu care trebuie să concorde clasamentul corespunzător scorurilor testului. În primul caz, se face presupunerea că factorul comun itemilor produce un clasament concordant celui dat de variabila latentă. În cel de-al doilea caz, nu se știe nimic despre clasamentul produs de variabila latentă dar se presupune

că toți itemii ale căror clasamente sunt concordante între ele produc clasamente concordante cu cel al variabilei latente.

15.3. VALIDITATEA UNUI TEST

15.3.1. Validitatea relativă la construct

Aprecierea validității unui test se bazează pe corectitudinea interpretărilor date scorurilor testului și a concluziilor desprinse din acestea. În majoritatea cazurilor, pentru interpretarea rezultatelor unei testări se compară scorurile fiecărei persoane examinate cu scorurile altor persoane. Compararea înseamnă să se identifice care dintre relațiile $>$, $=$ sau $<$ este prezentă între două scoruri.

Pentru a se putea face inferențe corecte despre persoane, plecând de la rezultatele unui test, este necesar ca relațiile $>$, $=$ sau $<$ existente între scorurile testului să reflecte corect relația existentă între valorile variabilei măsurate de test. S-a folosit aici expresia "să reflecte corect" în locul expresiei "să fie identică" deoarece, după cum se va arăta în continuare, identitatea relațiilor nu este, în general, posibilă. Relațiile dintre scorurile testului "reflectă corect" relațiile dintre valorile variabilei măsurate de test dacă ierarhia subiecților constituită pe baza scorurilor testului aproximează bine (într-un sens care va fi precizat în continuare) ierarhia existentă între aceste persoane în privința variabilei respective.

Notăm:

A = mulțimea persoanelor examinate cu un test;

X = mulțimea scorurilor posibile ale testului:

$X = \{x_1, x_2, \dots, x_n\}$, cu $x_1 < x_2 < \dots < x_n$

$t(a)$ = scorul la test pentru persoana $a \in A$;

$t(A) = \{t(a) \mid a \in A\}$

Presupunem că mulțimile A și X sunt finite.

Definim în mulțimea A trei relații, astfel:

$a \succ_T b \iff t(a) > t(b), a, b \in A$

$a =_T b \iff t(a) = t(b), a, b \in A$

$a \succeq_T b \iff a \succ_T b \text{ sau } a =_T b, a, b \in A$

Relația $=_T$ este o relație de echivalență, iar relația \succeq_T este o relație de preordine totală. Prin relația \succeq_T , testul realizează un clasament al persoanelor examinate.

În teoria răspunsului la itemi se consideră că variabila psihică măsurată de test poate lua orice valoare reală între $-\infty$ și $+\infty$.

Notăm :

$\theta(a)$ = valoarea variabilei psihice măsurate de test, pentru persoana $a \in A$;

$\theta(A) = \{\theta(a) | a \in A\}$

Întrucât valorile $\theta(a)$, $a \in A$, sunt numere reale, elementele mulțimii $\theta(A)$ pot fi comparate cu ajutorul relațiilor $>$, \geq și $=$ existente în \mathbb{R} . În mulțimea A vom defini trei relații binare astfel:

$a >_s b \Leftrightarrow \theta(a) > \theta(b)$, $a, b \in A$

$a =_s b \Leftrightarrow \theta(a) = \theta(b)$, $a, b \in A$

$a \geq_s b \Leftrightarrow a >_s b$ sau $a =_s b$, $a, b \in A$

Relația $=_s$ este o relație de echivalență, iar relația \geq_s este o relație de preordine totală.

Deci, în mulțimea A avem două clasamente, unul produs de relația \geq_T și unul produs de relația \geq_s .

Întrucât variabila psihică poate lua orice valoare reală, în timp ce mulțimea scorurilor posibile ale testului, X , este finită, în mod obișnuit numărul claselor de echivalență ale relației de preordine totală \geq_s este mai mare decât cel al claselor de echivalență în relația de preordine totală \geq_T .

Considerăm că un test care are funcția de măsurare a unei variabile psihice este un instrument de măsură valid dacă îndeplinește următoarele două condiții:

a. oricare două persoane care au aceeași valoare a variabilei psihice măsurate obțin la test același rezultat:

$$a =_s b \Rightarrow a =_T b$$

b. orice persoană poate fi superioară alteia la test numai dacă îi este superioară și în privința variabilei criteriu:

$$a >_T b \Rightarrow a >_s b$$

Aceste condiții conduc la următoarea formulare:

DEFINIȚIA 15.3.1.1. Un test este valid pentru mulțimea de persoane A atunci când clasamentele corespunzătoare relațiilor de preordine totală \geq_s și \geq_T coincid sau clasamentul corespunzător relației \geq_s este o rafinare a clasamentului corespunzător relației \geq_T .

Înseamnă că în cazul unui test valid în sensul acestei definiții oricare două persoane care au aceeași valoare pentru variabila psihică măsurată vor avea rezultate egale la test. Deci se poate construi o funcție

$$F: \mathbb{R} \rightarrow X$$

$F(\theta)$ = rezultatul obținut la test de toate persoanele care au valoarea θ pentru variabila psihică,

care are următoarea proprietate:

$$\theta_1 > \theta_2 \Rightarrow F(\theta_1) > F(\theta_2)$$

În realitate, însă, cu excepția testelor care au un singur rezultat posibil, ceea ce face să nu prezinte interes practic, nu se poate construi niciodată o asemenea funcție. Deci nici un test pentru care numărul scorurilor posibile este mai mare decât 1 nu satisface condițiile definiției 15.3.1.1. Aceasta, deoarece nici un test nu realizează o măsurare perfectă: din cauza a numeroși factori care intervin (condițiile de administrare și de cotare, starea de oboseală sau boală a celor examinați, hazardul etc.) pentru o valoare a variabilei psihice măsurate θ se pot obține mai multe rezultate de test. Mai exact, pentru fiecare valoare reală θ există o repartiție de probabilitate a scorurilor testului:

$$\begin{pmatrix} x_1 & x_2 & \dots & x_n \\ P_1(\theta) & P_2(\theta) & \dots & P_n(\theta) \end{pmatrix} \quad (15.3.1.1)$$

cu $0 \leq P_i(\theta) \leq 1$, $i=1,2,\dots,n$

și $P_1(\theta) + P_2(\theta) + \dots + P_n(\theta) = 1$,

unde $P_i(\theta)$ reprezintă probabilitatea scorului x_i , $i=1,2,\dots,n$.

Vom încerca să redefinim validitatea testului utilizând funcția care are ca valoare, în fiecare punct $\theta \in \mathbb{R}$, scorul cel mai probabil la test, $S_{\max}(\theta)$.

$$S_{\max}: \mathbb{R} \rightarrow X$$

$$S_{\max}(\theta) = \max\{x_j \in X | P_j(\theta) = \max\{P_i(\theta) | i=1,2,\dots,n\}\} \quad (15.3.1.2)$$

($S_{\max}(\theta)$ este valoarea de test care are cea mai mare probabilitate în punctul θ , dacă există mai multe valori pentru care probabilitatea este egală cu cea maximă, atunci $S_{\max}(\theta)$ este cea mai mare dintre ele.)

Considerăm că un test este un instrument de măsură bun numai dacă îndeplinește următoarele condiții:

(C1) două persoane care au aceeași valoare a variabilei psihice au șanse mari să obțină același rezultat la test;

(C2) există șanse mari ca dintre două persoane diferite în privința variabilei psihice, cea care are o valoare mai mare a acestei variabile să realizeze la test un rezultat cel puțin la fel de bun ca și cealaltă persoană.

Se observă că, dacă sunt îndeplinite condițiile:

• pentru fiecare valoare $\theta \in \mathbb{R}$:

• mulțimea $\{x_j \in X | P_j(\theta) = \max\{P_i(\theta) | i=1,2,\dots,n\}\}$ conține un singur element și

• probabilitatea rezultatului x_j , $P_j(\theta)$, este mult mai mare decât probabilitățile celorlalte scoruri de test în punctul θ ,

• funcția S_{max} este nedescrescătoare:
 $\theta_1 > \theta_2 \Rightarrow S_{max}(\theta_1) \geq S_{max}(\theta_2)$
 atunci sunt verificate condițiile (C1) și (C2) de mai sus. (15.3.1.3)

Atunci când în fiecare punct $\theta \in R$ există un singur scor cu probabilitate mare, iar această probabilitate este mult mai mare decât probabilitățile celorlalte scoruri, funcția S_{max} este apropiată de funcția S_{med} care ia ca valoare, în fiecare punct $\theta \in R$ scorul mediu al testului:

$$S_{med}: R \rightarrow [x_1, x_n]$$

$$S_{med}(t) = \sum_{i=1}^n x_i P_i(t) \quad (15.3.1.4)$$

Pentru măsurarea distanței dintre funcțiile S_{max} și S_{med} se poate folosi norma spațiului L_1 :

$$\|S_{max} - S_{med}\| = \int_{-\infty}^{+\infty} |S_{max}(t) - S_{med}(t)| dt \quad (15.3.1.5)$$

Cu ajutorul funcției S_{max} putem defini trei relații binare în mulțimea A:

$$a >_p b \Leftrightarrow S_{max}(\theta(a)) > S_{max}(\theta(b)), \quad a, b \in A$$

$$a =_p b \Leftrightarrow S_{max}(\theta(a)) = S_{max}(\theta(b)), \quad a, b \in A$$

$$a \geq_p b \Leftrightarrow a >_p b \text{ sau } a =_p b, \quad a, b \in A$$

Relația $=_p$ este o relație de echivalență, iar relația $>_p$ este o relație de preordine totală.

Dacă este îndeplinită condiția (15.3.1.3) atunci au loc implicațiile:

$$a =_s b \Leftrightarrow \theta(a) = \theta(b) \Rightarrow a =_p b$$

$$a >_p b \Leftrightarrow S_{max}(\theta(a)) > S_{max}(\theta(b)) \Rightarrow \theta(a) > \theta(b) \Leftrightarrow a >_s b$$

Se deduce că, dacă este îndeplinită condiția (15.3.1.3) atunci clasamentul corespunzător relației \geq_s coincide cu clasamentul corespunzător relației \geq_p sau este o rafinare a acestuia.

DEFINIȚIA 15.3.1.2. Vom spune că un test este P-valid pentru mulțimea persoanelor examinate A atunci când clasamentele corespunzătoare relațiilor de preordine totală \geq_s și \geq_p coincid sau clasamentul corespunzător relației \geq_s este o rafinare a clasamentului corespunzător relației \geq_p .

Condiția de P-validitate este mai slabă decât condiția de validitate formulată prin definiția 15.3.1.1. Dacă un test este valid, atunci pentru orice valoare $\theta \in R$ există un singur scor de test cu probabilitatea diferită de zero, care

corespunde valorii funcției F în punctul θ , deci funcția S_{max} coincide cu funcția F și este verificată condiția (15.3.1.3), adică testul este P-valid. Dar nu orice test P-valid este valid.

Se pune, în mod firesc, întrebarea: cum se poate construi un test P-valid? Având o mulțime de itemi M, ale căror funcții caracteristice $p_i, i=1, 2, \dots, n$, sunt cunoscute, care itemi ar trebui aleși din M pentru a rezulta un test P-valid?

Se fac următoarele convenții, obișnuite în teoria răspunsului la itemi:

a. toți itemii testului măsoară aceeași variabilă psihică;

b. la orice valoare $\theta \in R$ fixată, modul de rezolvare a unui item este independent de modul de rezolvare a celorlalți itemi ai testului.

Au fost determinate următoarele condiții referitoare la funcțiile caracteristice ale itemilor, care asigură obținerea unui test P-valid.

TEOREMA 15.3.1.1. (Albu, Pitariu, 1991a). Dacă funcțiile $p_i, i=1, \dots, n$ sunt derivabile și nedescrescătoare pe R, $p_i(-\infty)=0, p_i(+\infty)=1$ pentru $i=1, 2, \dots, n$ atunci există $k+1$ puncte ($0 < k < n-1$):

$$-\infty < t_0 < t_1 < \dots < t_k < +\infty$$

astfel încât:

$$S_{max}(\theta) = \begin{cases} 0, & 0 < t_0 \\ i_1, & 0 < \theta < t_1 \\ i_2, & \theta \in [t_1, t_2) \\ \dots \\ i_k, & \theta \in [t_{k-1}, t_k) \\ n, & \theta \geq t_k \end{cases} \quad (15.3.1.6)$$

cu $0 < i_1 < i_2 < \dots < i_k < n$.

Deci, în aceste condiții, funcția S_{max} este o funcție în scară, nedescrescătoare. Lungimile "treptelor" scării (adică a segmentelor care corespund câte unei valori de scor) sunt determinate tot de funcțiile p_i . Prezentăm, în acest sens, următoarea teoremă.

TEOREMA 15.3.1.2. (Albu, Pitariu, 1992a). Dacă funcțiile $p_i, i=1, 2, \dots, n$, sunt continue pe R și dacă există punctele

$$T_{1,min} < T_{1,max} \leq T_{2,min} < T_{2,max} \leq \dots \leq T_{n,min} < T_{n,max}$$

astfel încât

$$p_i(\theta) \begin{cases} =0, & \theta \in (-\infty, T_{i,\min}] \\ \in(0,1), & \theta \in (T_{i,\min}, T_{i,\max}) \\ =1, & \theta \in [T_{i,\max}, +\infty) \end{cases}$$

pentru $i=1,2,\dots,n$, fiecare funcție p_i fiind crescătoare pe intervalul $(T_{i,\min}, T_{i,\max})$, atunci există n puncte $t_i \in (T_{i,\min}, T_{i,\max})$, $i=1,2,\dots,n$, cu proprietatea că

$$S_{\max}(\theta) = \begin{cases} 0, & \theta < t_1 \\ k, & \theta \in [t_k, t_{k+1}), k=1,2,\dots,n-1 \\ n, & \theta \geq t_n \end{cases}$$

Punctele t_i sunt soluțiile ecuațiilor $p_i(\theta) = 0.5$ pentru $i=1,2,\dots,n$.

Condițiile impuse de teorema 15.3.1.2 asupra funcțiilor p_i și graficul funcției S_{\max} corespunzătoare sunt ilustrate în figura 15.3.1.1.

Pentru un test ai cărui itemi satisfac ipotezele teoremei 15.3.1.2 are loc egalitatea:

$$\|S_{\max} - S_{\text{med}}\| = \sum_{k=1}^n \left\{ \int_{T_{k,\min}}^{t_k} p_k(t) dt + \int_{t_k}^{T_{k,\max}} (1-p_k(t)) dt \right\}$$

(15.3.1.7)

Această cantitate este reprezentată de suma ariilor suprafețelor hașurate din Figura 15.3.1.1.

Pentru minimizarea valorii $\|S_{\max} - S_{\text{med}}\|$ se vor alege itemi ale căror funcții caracteristice au intervale $[T_{k,\min}, T_{k,\max}]$ cu lungimi mici.

Un test ai cărui itemi satisfac ipotezele teoremei 15.3.1.2 prezintă încă un avantaj important: pe fiecare interval $I_k = \{\theta \in \mathbb{R} | S_{\max}(\theta) = k\}$, $k=1,2,\dots,n-1$, pe lângă scorul k , care are probabilitatea maximă, au probabilitate diferită de zero doar scorurile $k-1$ și $k+1$.

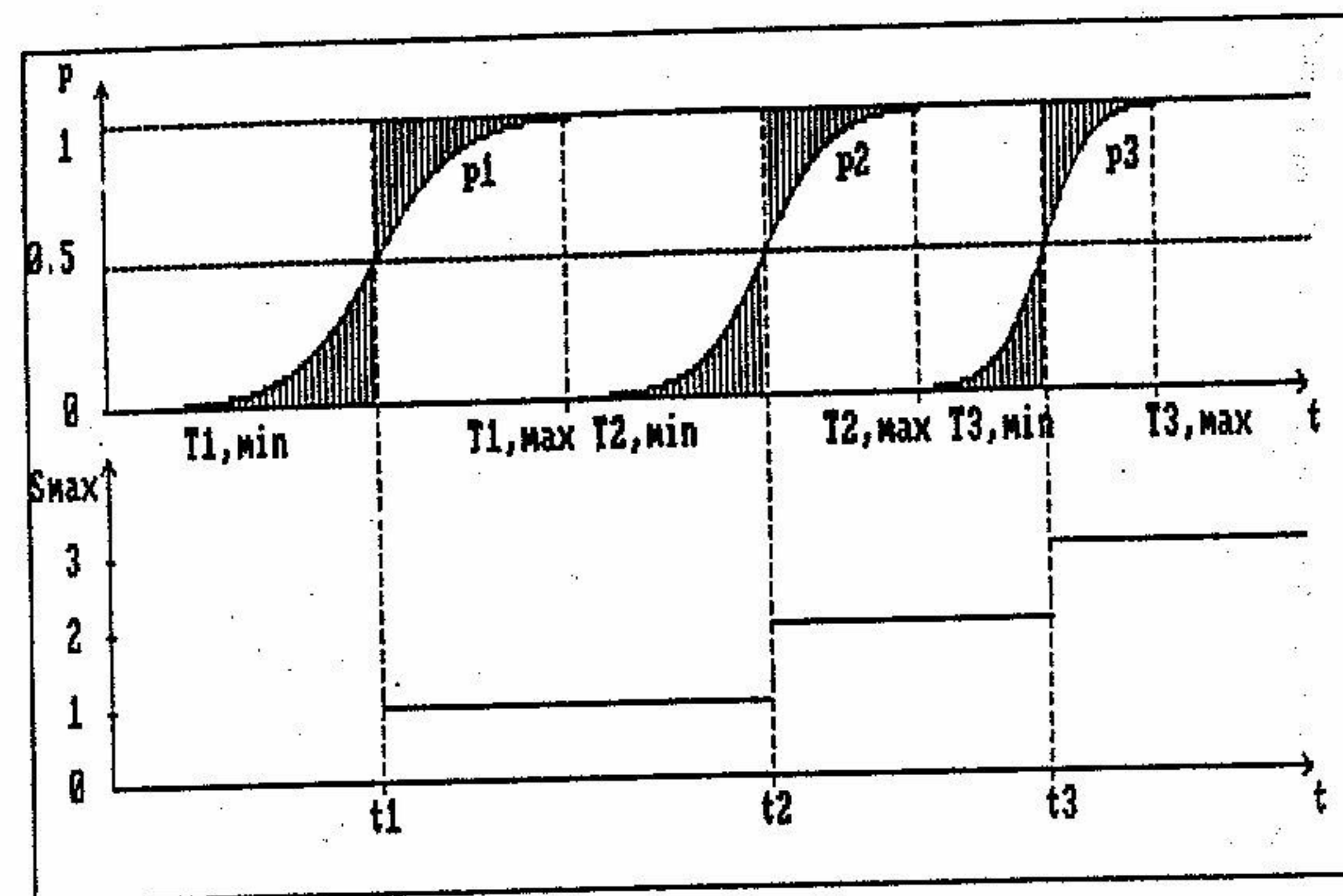


Figura 15.3.1.1

Legătura dintre funcția S_{\max} și funcțiile caracteristice ale itemilor

15.3.2. Validitatea relativă la criteriu

Validarea unui test relativă la un criteriu constă în compararea ierarhiei produse de scorurile testului cu o altă ierarhie, realizată de o variabilă criteriu. Adică, validarea relativă la criteriu constă în compararea a două clasamente.

Notăm:

A = o mulțime formată din N persoane care au fost examinate cu testul a cărui validitate se studiază și pentru care se cunosc valorile variabilei criteriu;

T = mulțimea scorurilor de test obținute de persoanele din mulțimea A:

$T = \{t_1, t_2, \dots, t_n\}$, $t_i \in \mathbb{R}$ pentru $i=1,2,\dots,n$ și $t_1 < t_2 < \dots < t_n$

C = mulțimea valorilor criteriului pentru persoanele din mulțimea A;

$C = \{c_1, c_2, \dots, c_m\}$, $c_j \in \mathbb{R}$ pentru $j=1,2,\dots,m$ și $c_1 < c_2 < \dots < c_m$

$t(a)$ = scorul la test al unei persoane $a \in A$;

$c(a)$ = valoarea criteriului pentru persoana $a \in A$;

Definim în mulțimea A următoarele relații:

pentru orice $a, b \in A$:

$$a >_T b \Leftrightarrow I(a) > I(b)$$

$$a =_T b \Leftrightarrow I(a) = I(b)$$

$$a \geq_T b \Leftrightarrow a >_T b \text{ sau } a =_T b$$

$$a >_C b \Leftrightarrow c(a) > c(b)$$

$$a =_C b \Leftrightarrow c(a) = c(b)$$

$$a \geq_C b \Leftrightarrow a >_C b \text{ sau } a =_C b$$

Relațiile \geq_T și \geq_C sunt preordini totale în A.

Considerăm că un test măsoară același lucru ca și criteriul, adică este valid relativ la criteriu, dacă îndeplinește următoarele două condiții:

(D1) două persoane care au rezultate egale la criteriu obțin rezultate egale și la test:

$$a =_C b \Rightarrow a =_T b \quad a, b \in A$$

(D2) dacă o persoană îi este superioară alteia la criteriu, atunci îi este superioară și la test:

$$a >_C b \Rightarrow a >_T b \quad a, b \in A$$

Cele două condiții impun ca în mulțimea A clasamentele \geq_T și \geq_C să fie identice:

$$a \geq_T b \Leftrightarrow a \geq_C b, \quad a, b \in A.$$

Dar acest lucru înseamnă că numărul claselor de echivalență în raport cu relațiile $=_T$ și $=_C$ este același. Fiecare clasă de echivalență pentru relația $=_T$ corespunde unei valori de test, iar fiecare clasă de echivalență pentru relația $=_C$ corespunde unei valori a criteriului. Or, în multe situații întâlnite în practică, numărul valorilor posibile ale testului diferă de numărul valorilor posibile ale criteriului, ceea ce face ca să nu poată fi îndeplinite condițiile (D1) și (D2). Din acest motiv, am înlocuit aceste condiții cu următoarele:

• dacă $n < m$ atunci:

$$(D1.1) \quad a =_C b \Rightarrow a =_T b$$

$$(D2.1) \quad a >_C b \Rightarrow a \geq_T b$$

• dacă $n > m$, atunci:

$$(D1.2) \quad a =_T b \Rightarrow a =_C b$$

$$(D2.2) \quad a >_T b \Rightarrow a \geq_C b$$

• dacă $n = m$, atunci:

$$(D1.3) \quad a =_C b \Rightarrow a =_T b$$

$$(D2.3) \quad a >_C b \Rightarrow a >_T b$$

Aceste condiții se pot transcrie astfel:

- dacă $n < m$, atunci partiția ordonată corespunzătoare relației \geq_C este o rafinare a partiției ordonate corespunzătoare relației \geq_T ;
- dacă $n > m$, atunci partiția ordonată corespunzătoare relației \geq_T este o rafinare a partiției ordonate corespunzătoare relației \geq_C ;
- dacă $n = m$, atunci relațiile \geq_T și \geq_C sunt identice.

Ele conduc la următoarea definiție:

DEFINIȚIA 15.3.2.1. Un test este valid relativ la un criteriu pentru mulțimea de persoane A dacă partițiile ordonate corespunzătoare relațiilor de preordine \geq_T și \geq_C coincid sau una este o rafinare a celeilalte.

Se observă că atunci când coeficientul de validitate se exprimă prin coeficientul de corelație a rangurilor al lui Spearman, el ia valoarea 1 dacă și numai dacă cele două clasamente, \geq_C și \geq_T , coincid. Dacă, însă, coeficientul de validitate este dat de coeficientul de corelație liniară între test și criteriu, atunci sunt adevărate următoarele afirmații:

• obținerea valorii 1 a coeficientului de validitate înseamnă că sunt verificate condițiile din definiția 15.3.2.1;

• nu întotdeauna când sunt satisfăcute condițiile definiției 15.3.2.1 coeficientul de validitate ia valoarea 1.

Deci coeficientul de corelație liniară nu ne permite să deducem dacă un test este valid relativ la criteriu în sensul definiției 15.3.2.1.

În continuare propunem o metodă de analiză a relației test-criteriu, utilă pentru aprecierea validității relative la criteriu a unui test atunci când mulțimea A conține un număr mare de persoane.

Pe baza valorilor variabilei criteriu și a valorilor obținute la test de către subiecții din mulțimea A se alcătuește următorul tabel de frecvențe:

Test	Criteriu		
	c_1	\dots	$c_j \dots c_m$
t_1	$f_{1,1}$	\dots	$f_{1,j} \dots f_{1,m}$
\dots	\dots	\dots	\dots
t_i	$f_{i,1}$	\dots	$f_{i,j} \dots f_{i,m}$
\dots	\dots	\dots	\dots
t_n	$f_{n,1}$	\dots	$f_{n,j} \dots f_{n,m}$

unde f_{ij} = numărul de subiecți care au valoarea c_j la criteriu și scorul t_i la test, $i=1,2,\dots,n$; $j=1,2,\dots,m$.

Din felul în care au fost definite mulțimile C și T rezultă:

$$\sum_{i=1}^n \sum_{j=1}^m f_{ij} = N$$

$$\sum_{i=1}^n f_{ij} > 0, \quad j=1,2,\dots,m$$

$$\sum_{j=1}^m f_{ij} > 0, \quad i=1,2,\dots,n$$

Pentru fiecare pereche de valori de test $t_i, t_j \in T$, $i < j$, se calculează probabilitatea ca un subiect care a obținut rezultatul t_i să fie inferior, în privința variabilei criteriu, unei persoane care a realizat la test valoarea t_j :

$$P_{ij} = P(a <_c b \mid t(b)=t_i \text{ și } t(a)=t_j) \quad (15.3.2.1)$$

$$P_{ij} = \frac{\sum_{k=2}^m \sum_{h=1}^{k-1} f_{j,k} f_{i,h}}{\left(\sum_{k=1}^m f_{j,k}\right) \left(\sum_{h=1}^m f_{i,h}\right)} \quad (15.3.2.2)$$

Se definește $P_{ji} = P_{ij}$, $i, j=1,2,\dots,n$.

Pentru fiecare valoare de test $t_i \in T$ se calculează probabilitatea ca doi subiecți care au obținut scorul t_i să aibă valori egale la variabila criteriu:

$$P_{ii} = P(a =_c b \mid t(a)=t_i \text{ și } t(b)=t_i) \quad (15.3.2.3)$$

$$P_{ii} = \begin{cases} \frac{\sum_{k=1}^m f_{i,k}(f_{i,k}-1)}{\left(\sum_{k=1}^m f_{i,k}\right)\left(\sum_{k=1}^m f_{i,k}-1\right)}, & \text{daca } \sum_{k=1}^m f_{i,k} > 1 \\ 1, & \text{daca } \sum_{k=1}^m f_{i,k} = 1 \end{cases} \quad (15.3.2.4)$$

Se obține astfel o matrice simetrică $B=(P_{ij})_{i,j=1,\dots,n}$

Probabilitățile P_{ij} care se abat cel mai mult de la valoarea 1 corespund perechilor de valori de test care nu pot fi interpretate corect. Atunci când $i < j$ și $P_{ij} < 0.5$, este foarte riscant să se afirme despre un subiect care a obținut la test rezultatul t_i că este superior unei persoane care a realizat valoarea t_j .

Definim doi coeficienți de validitate relativă la criteriu, prin următoarele formule (Albu, Pitariu, 1994):

$$V_{dif} = \min \{P_{ij} \mid i, j=1,2,\dots,n \text{ și } i \neq j\} \quad (15.3.2.5)$$

$$V_{eg} = \min \{P_{ii} \mid i=1,2,\dots,n\} \quad (15.3.2.6)$$

Aceștia se interpretează astfel:

- Există o probabilitate de cel puțin $100V_{dif}\%$ ca un subiect a care a obținut la test un rezultat mai bun decât un alt subiect b ($a >_T b$), să fie mai bun decât acesta și în privința variabilei criteriu ($a >_c b$).

- Există o probabilitate de cel puțin $100V_{eg}\%$ ca doi subiecți, a și b , care au realizat aceeași valoare la test ($a =_T b$) să fie egali și la variabila criteriu ($a =_c b$).

Legătura dintre validitatea relativă la criteriu a testului și coeficienții V_{eg} și V_{dif} poate fi formulată astfel:

TEOREMA 15.3.2.1

- Când $m=n$ atunci testul este valid dacă și numai dacă $V_{eg}=V_{dif}=1$.
- Dacă $m < n$ și testul este valid atunci $V_{eg}=1$.
- Dacă $m > n$ și testul este valid atunci $V_{dif}=1$.

Se poate îmbunătăți valoarea coeficientului V_{dif} pentru un test ținând cont de următoarea observație: Dacă se înlocuiesc două valori de test consecutive, t_i și t_{i+1} , cu o valoare nouă, $t_{i \& i+1}$, atunci fiecare probabilitate $P_{i \& i+1, k}$ va fi cuprinsă în intervalul

$$(\min\{P_{ik}, P_{i+1,k}\}, \max\{P_{ik}, P_{i+1,k}\}),$$

$k=1,2,\dots,n$.

Deci se poate mări valoarea coeficientului V_{dif} prin reunirea a două sau mai multe valori de test consecutive.

Înlocuind valorile t_1, t_2, \dots, t_k , $k \geq 1$, din T cu o valoare nouă, t , cardinalul mulțimii T , n , scade. Coeficientul V_{dif} crește, dar este posibil ca valoarea coeficientului V_{eg} să scadă.

Exemplu.

S-a aplicat acest procedeu pentru o scală nouă, A , alcătuită din itemi extrași din Inventarul Psihologic California (CPI). Drept variabilă criteriu s-a folosit o scală existentă în CPI, scala Dominanță (Do).

În figura 15.3.2.1 sunt reprezentate sub forma unei dendrograme grupările de scoruri realizate de algoritm.

În tabelul 15.3.2.1 sunt notate, pentru fiecare pas, valorile coeficienților V_{df} și V_{cr} . Se constată că, în timp ce coeficientul V_{df} crește de la 0.442, deci de la o valoare necorespunzătoare pentru test, la 0.793, coeficientul V_{cr} oscilează.

Se impune o observație: Coeficientul de corelație dintre scalele A și Do este egal cu 0.856, pentru $N=1065$. În conformitate cu procedeul obișnuit de lucru, s-ar deduce că scala A are o validitate bună relativ la scala Do. Or, probabilitatea de a interpreta corect rezultatele obținute de doi subiecți este, pentru unele perechi de scoruri, mai mică decât 50%.

Tabelul 15.3.2.1
Valorile coeficienților de validitate
la fiecare pas al algoritmului

Grupele de scoruri formate	V_{df}	V_{cr}
-	0.442	0.076
14,15	0.511	0.076
2,3	0.521	0.076
9,10	0.532	0.076
6,7	0.535	0.076
12,13	0.540	0.076
4,5	0.576	0.076
8,(9,10)	0.639	0.076
(12,13),(14,15)	0.645	0.076
0,1	0.664	0.082
(4,5),(6,7)	0.683	0.082
(0,1),(2,3)	0.689	0.079
(8,(9,10)),11	0.786	0.079
((12,13),(14,15)),16	0.793	0.079

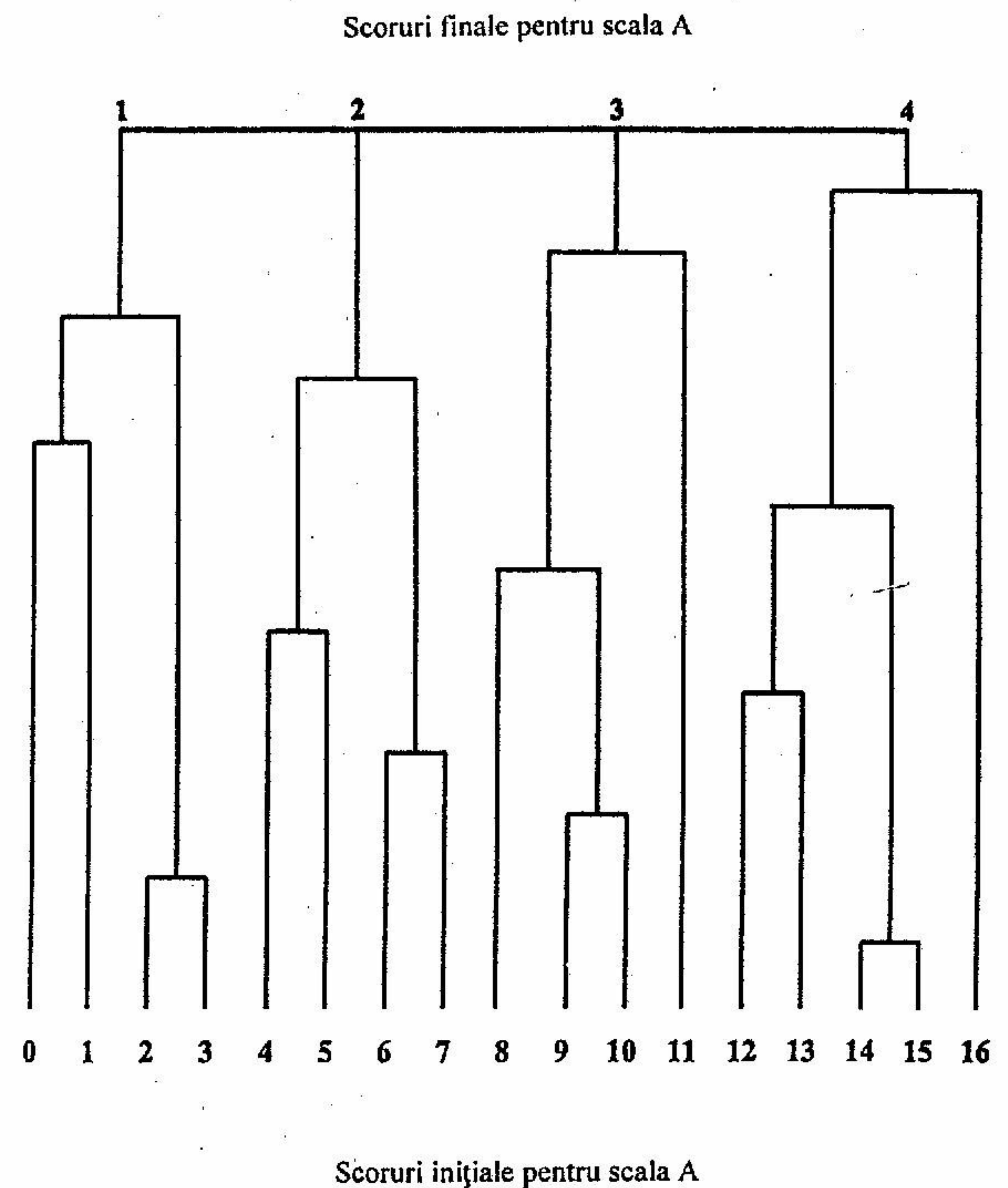


Figura 15.3.2.1
Gruparea scorurilor pentru scala A

ANEXA I

NOȚIUNI DE TEORIA CLASIFICĂRII

I.1. RELAȚII CARE INTERVIN ÎN CLASIFICARE

Fie A o mulțime finită și R o relație binară definită pe A . Pentru $a, b \in A$ notăm $\text{non}(aRb)$ dacă nu are loc aRb .

DEFINIȚIA I.1.1.

R este reflexivă $\Leftrightarrow \forall a \in A: aRa$

(orice element este în relația R cu el însuși);

R este simetrică $\Leftrightarrow \forall a, b \in A: aRb \Rightarrow bRa$

(dacă un element a este în relația R cu un element b , atunci și elementul b este în relația R cu a);

R este antisimetrică $\Leftrightarrow \forall a, b \in A, a \neq b: aRb \Rightarrow \text{non}(bRa)$

(dacă un element a este în relația R cu un element b distinct de a , atunci elementul b nu este în relația R cu a);

R este tranzitivă $\Leftrightarrow \forall a, b, c \in A: aRb \text{ și } bRc \Rightarrow aRc$;

(pentru oricare trei elemente, a , b și c dacă a este în relația R cu b , iar b este în relația R cu c , atunci a este în relația R și cu c);

R este completă $\Leftrightarrow \forall a, b \in A \Rightarrow aRb \text{ sau } bRa$.

(pentru orice pereche de elemente, a și b , distincte sau nu între ele, are loc cel puțin una dintre următoarele două afirmații: " a este în relația R cu b " și " b este în relația R cu a).

Exemplul I.1.1.

Considerăm mulțimea A formată din elevii unei clase care au fost examinați cu un test psihologic. Pentru fiecare elev, $a \in A$, notăm cu $s(a)$ scorul obținut la test.

În mulțimea A definim trei relații binare, S , E și C , astfel: pentru oricare două elemente $a, b \in A$:

$$aSb \Leftrightarrow s(a) > s(b)$$

(a este în relația S cu b dacă a este "superior" lui b la test, adică a a obținut un scor mai mare decât b);

$$aEb \Leftrightarrow s(a) = s(b)$$

(a este în relația E cu b dacă a este "egal" cu b la test, adică a și b au obținut același scor);

$$aCb \Leftrightarrow s(a) \geq s(b)$$

(a este în relația C cu b dacă a este "cel puțin la fel de bun ca și b la test, adică a a obținut un scor egal cu cel al lui b sau mai mare decât acesta).

Aceste relații au următoarele proprietăți:

- Pentru orice element a din A , are loc $s(a) = s(a)$, adică aEa . Prin urmare, relația E este reflexivă.

- Analog se arată că relația C este reflexivă.

- Relația S nu este reflexivă, pentru că nici un elev nu a obținut scor mai mare decât scorul propriu: nu are loc $s(a) > s(a)$, pentru nici un element a din A .

- Pentru oricare două elemente din A , a și b , dacă scorul lui a este egal cu scorul lui b atunci și scorul lui b este egal cu cel al lui a , deci, din aEb rezultă bEa . Prin urmare, relația E este simetrică.

- Dacă scorul unui element a este mai mare decât scorul unui element b , atunci scorul lui b nu poate fi mai mare decât scorul lui a . Deci, din aSb rezultă $\text{non}(bSa)$. Prin urmare, relația S este antisimetrică.

- Dacă între scorurile a doi elevi, a și b , există inegalitatea $s(a) \geq s(b)$, nu se poate ști dacă are loc și inegalitatea $s(b) \geq s(a)$ (care ar fi adevărată dacă $s(a) = s(b)$) sau are loc $\text{non}(s(b) \geq s(a))$ (ceea ce ar fi adevărat dacă $s(a) > s(b)$). Prin urmare, relația C nu este nici simetrică, nici antisimetrică.

- Dacă un element a din A are un scor egal cu cel al unui element b , care este egal cu cel al unui element c , atunci scorul lui a este egal cu cel al lui c . Adică, din aEb și bEc rezultă aEc . Prin urmare, relația E este tranzitivă.

- Analog se arată că relațiile S și C sunt tranzitive.

- Pentru oricare două elemente din A , a și b , este adevărată una din relațiile $s(a) > s(b)$, $s(a) = s(b)$, $s(b) > s(a)$. Înseamnă că are loc cel puțin una dintre inegalitățile $s(a) \geq s(b)$ și $s(b) \geq s(a)$, respectiv este adevărată cel puțin una dintre afirmațiile " aCb " și " bCa ". Adică relația C este completă.

- Relațiile S și E nu sunt complete.

DEFINIȚIA I.1.2. O relație reflexivă, tranzitivă și simetrică se numește relație de echivalență.

Dacă R este o relație de echivalență și pentru două elemente $a, b \in A$ are loc aRb , se spune că " a și b sunt echivalente". Fiind dată o relație de echivalență pe mulțimea A , orice submulțime a lui A formată din elemente echivalente între ele formează o clasă de echivalență.

Exemplul I.1.2.

Relația E din exemplul I.1.1. este o relație de echivalență.

O clasă de echivalență este formată din toți elevii care au obținut același scor la test.

DEFINIȚIA I.1.3. O relație tranzitivă și completă se numește preordine totală.

Exemplul I.1.3.

Relația C este o preordine totală.

Dacă R este o preordine totală și pentru două elemente a și b din mulțimea A are loc aRb vom spune că " a este preferat lui b ".

DEFINIȚIA I.1.4. O preordine totală antisimetrică se numește ordine totală

Exemplul I.1.4.

Dacă în mulțimea A nu există elevi care să fi obținut același scor, atunci relația S este o ordine totală.

TEOREMA I.1.1. Dacă R este o relație de preordine totală pe A , atunci relația I , definită pe A astfel:

pentru orice $a, b \in A$, $aIb \Leftrightarrow aRb$ și bRa (I.1.1)
este o relație de echivalență.

Exemplul I.1.5.

Utilizând relațiile definite în exemplul I.1.1 se constată că relația I definită prin formula (I.1.1) cu ajutorul relației C

$aIb \Leftrightarrow aCb$ și bCa

coincide cu relația E :

pentru orice $a, b \in A$, $aIb \Leftrightarrow s(a) \geq s(b)$ și $s(b) \geq s(a) \Leftrightarrow s(a) = s(b) \Leftrightarrow aEb$

I.2. PARTIȚII

Fie E o mulțime finită cu n elemente.

DEFINIȚIA I.2.1. O partiție a mulțimii E , $P(E)$, este o mulțime de părți ale lui E :

$P(E) = \{E_1, E_2, \dots, E_k\}$, $E_i \subseteq E$, $i=1, 2, \dots, k$
două câte două disjuncte:

$E_i \cap E_j = \emptyset$ pentru orice $i \neq j$ (I.2.1)
și a căror reuniune este egală cu E :

$$E_1 \cup E_2 \cup \dots \cup E_k = E$$

(I.2.2)

Aceste părți se numesc clasele partiției.

Indicarea unei partiții pe o mulțime E este același lucru cu indicarea unei relații de echivalență pe E : două elemente $x, y \in E$ sunt echivalente dacă și numai dacă x și y aparțin aceleiași clase a partiției.

Dacă pe mulțimea E a fost definită o relație de preordine totală R , atunci relația de echivalență I definită prin formula (I.1.1) determină o partiție a mulțimii E . Clasele partiției au proprietatea că ele pot fi așezate în ordine, astfel încât, pentru orice clasă aleasă, E_i , fiecare element al ei, $a \in E_i$, este în relațiile aRb și $\text{non}(bRa)$ (I.2.3)

cu fiecare element b al oricărei clase E_j aflate în dreapta clasei E_i .

Exemplul I.2.1.

Să reluăm exemplul I.1.1 și să notăm cu s_1, s_2, \dots, s_k scorurile care au fost obținute de elevii mulțimii A , așezate în ordine descrescătoare: $s_1 > s_2 > \dots > s_k$.

Notăm cu A_i mulțimea elevilor din A care au obținut scorul s_i , $i=1, 2, \dots, k$.

Mulțimile A_1, A_2, \dots, A_k formează o partiție a mulțimii A . Orice element a al unei clase A_i este în relația aCb și $\text{non}(bCa)$, adică $s(a) > s(b)$, cu orice element b al claselor $A_{i+1}, A_{i+2}, \dots, A_k$, pentru $i=1, 2, \dots, k-1$.

DEFINIȚIA I.2.2. Un vector cu componente mulțimi, (E_1, E_2, \dots, E_k) , $E_i \subseteq E$, $i=1, 2, \dots, k$, este o partiție ordonată pe mulțimea E , corespunzătoare relației de preordine totală R dacă îndeplinește următoarele condiții:

- pentru orice mulțime E_i , $i=1, 2, \dots, k$, și oricare două elemente $a, b \in E_i$ au loc relațiile aRb și bRa ;
- pentru oricare două mulțimi, E_i și E_j , cu $i < j$, și oricare două elemente $a \in E_i$ și $b \in E_j$, au loc relațiile aRb și $\text{non}(bRa)$.

Oricărei relații de preordine totală pe mulțimea E , R , îi corespunde o singură partiție ordonată pe E .

Fie

$$(A_1, A_2, \dots, A_m)$$

(I.2.4)

și

$$(B_1, B_2, \dots, B_p)$$

(I.2.5)

două partiții ordonate pe mulțimea E , corespunzătoare la două relații de preordine totală, R_1 și R_2 .

DEFINIȚIA 1.2.3. Se spune că partițiile ordonate (I.2.4) și (I.2.5) **coincid** dacă $m=p$ și pentru orice $i=1,2,\dots,m$ are loc $A_i=B_i$.

DEFINIȚIA 1.2.4. Se spune că partiția ordonată (I.2.4) este o **rafinare** a partiției ordonate (I.2.5) dacă $m>p$ și orice clasă de echivalență $A_i, i=1,2,\dots,m$, este o submulțime a unei clase de echivalență $B_j, j \in \{1,2,\dots,p\}$: $A_i \subseteq B_j$.

I.3. NUMEROTARE A UNEI PREORDINI TOTALE

Fie E o mulțime finită cu n elemente și R o relație de preordine totală pe E .

DEFINIȚIA 1.3.1. Se numește **numerotare** a elementelor lui E o funcție bijectivă $f: E \rightarrow \{1,2,\dots,n\}$.

Prin urmare, numerotarea constă în atribuirea de numere consecutive elementelor din E , începând cu 1, astfel încât oricare două elemente să primească numere diferite.

Convenim să notăm cu e_i elementul din E pentru care $f(e_i)=i, i=1,2,\dots,n$.

DEFINIȚIA 1.3.2. O numerotare a mulțimii $E, f: E \rightarrow \{1,2,\dots,n\}$ este o **numerotare a preordinii totale** R dacă pentru oricare două elemente distincte $a,b \in E$ are loc implicația:

$$f(a) < f(b) \Rightarrow aRb \quad (I.3.1)$$

Aceasta înseamnă că sunt adevărate relațiile $e_i R e_j$ pentru orice $i=1,2,\dots,n-1$ și $j=i+1,i+2,\dots,n$.

Observație. Din faptul că pentru două elemente distincte $a,b \in E$ are loc aRb nu se deduce $f(a) < f(b)$, întrucât este posibil ca împreună cu relația aRb să fie adevărată și relația bRa .

TEOREMA 1.3.1. Pentru orice relație de preordine totală R există cel puțin o numerotare a preordinii.

Demonstrația acestei teoreme este constructivă, indicând metoda de definire a funcției f . Se procedează conform algoritmului următor:

(1) Se formează partiția ordonată corespunzătoare relației R pe mulțimea E :

$$(E_1, E_2, \dots, E_m)$$

(2) Pentru fiecare clasă a partiției, $E_i, i=1,2,\dots,m$, se notează cu n_i numărul de elemente componente. Evident, $n_1+n_2+\dots+n_m=n$.

(3) Se definește o funcție bijectivă $f: E \rightarrow \{1,2,\dots,n\}$ astfel:

$$\forall e \in E_1: f(e) \in \{1,2,\dots,n_1\}$$

$$\forall e \in E_2: f(e) \in \{n_1+1, n_1+2, \dots, n_1+n_2\}$$

...

$$\forall e \in E_m: f(e) \in \{n_1+n_2+\dots+n_{m-1}+1, n_1+n_2+\dots+n_{m-1}+2, \dots, n_1+n_2+\dots+n_m\}$$

Observație.

Cu excepția cazului când fiecare clasă a partiției conține un singur element, pentru preordinea totală R pot fi definite mai multe numerotări. Dar, orice numerotare f are proprietatea că media valorilor sale pe o clasă E_i este egală cu:

$$m(E_i) = n_{i-1} + \frac{n_i+1}{2} \quad (I.3.2)$$

$i=1,2,\dots,m$, cu convenția $n_0=0$.

DEFINIȚIA 1.3.3. Funcția $r: E \rightarrow R$ definită prin:

$$\forall e \in E_i: r(e) = m(E_i), \quad i=1,2,\dots,m \quad (I.3.3)$$

se numește **funcție rang**. Valoarea ei pentru un element $e \in E$ se numește **rangul elementului** e în preordinea totală R .

Toate elementele din E care aparțin aceleiași clase de echivalență au același rang. Numărul valorilor distincte ale funcției rang este egal cu numărul claselor de echivalență ale partiției ordonate determinată de relația de preordine R .

Exemplul 1.3.1.

Fie $E = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7\}$ o mulțime de elevi care au fost examinați cu un test de atenție. În următorul tabel este notat numărul greșelilor făcute de fiecare elev:

Elev	e_1	e_2	e_3	e_4	e_5	e_6	e_7
Nr. erori	10	11	9	10	8	12	11

Pe mulțimea E se definește o relație R , astfel:

$$e_i R e_j \Leftrightarrow e_i \text{ a făcut mai puține erori sau tot atâtea ca și } e_j, \quad i=1,2,\dots,7.$$

Relația R este o relație de preordine totală pe mulțimea elevilor. Partiția ordonată corespunzătoare ei este:

$(\{e_5\}, \{e_3\}, \{e_1, e_4\}, \{e_2, e_7\}, \{e_6\})$.

Valorile funcției rang $r: E \rightarrow \mathbf{R}$ se calculează astfel:

$$r(e_5) = 0 + (1+1)/2 = 1$$

$$r(e_3) = 1 + (1+1)/2 = 2$$

$$r(e_1) = r(e_4) = 2 + (2+1)/2 = 3.5$$

$$r(e_2) = r(e_7) = 4 + (2+1)/2 = 5.5$$

$$r(e_6) = 6 + (1+1)/2 = 7$$

I.4. CLASAMENTE

Fie E o mulțime finită cu n elemente, R o relație de preordine totală pe E și r funcția rang asociată relației R .

DEFINIȚIA 1.4.1. Se numește **clasament** sau **ierarhie** pe mulțimea E ordonarea elementelor mulțimii E stabilită de funcția rang r , în care locul fiecărui element este egal cu rangul său.

Deci, un clasament este totdeauna asociat unei relații de preordine totală.

Exemplul 1.4.1. Pentru elevii mulțimii E din exemplul 1.3.1 clasamentul este următorul:

Elev	Loc
e_5	1
e_3	2
e_1	3.5
e_4	3.5
e_2	5.5
e_7	5.5
e_6	7

Să considerăm două relații de preordine totală pe mulțimea E , R_1 și R_2 , și funcțiile rang asociate acestora, r_1 și r_2 .

Pentru fiecare element $e \in E$ notăm cu $r_1(e)$ locul ocupat în primul clasament și cu $r_2(e)$ locul ocupat în al doilea clasament, $r_1(e)$ și $r_2(e)$ sunt rangurile elementului e în cele două preordini totale.

DEFINIȚIA 1.4.2. Clasamentele asociate relațiilor de preordine totală R_1 și R_2 coincid dacă pentru orice element $e \in E$ are loc egalitatea $r_1(e) = r_2(e)$.

TEOREMA 1.4.1. Clasamentele asociate relațiilor de preordine totală R_1 și R_2 coincid dacă și numai dacă relațiile R_1 și R_2 sunt identice.

DEFINIȚIA 1.4.3. Clasamentul asociat relației de preordine totală R_1 este o rafinare a clasamentului asociat relației de preordine totală R_2 dacă sunt îndeplinite următoarele condiții:

- pentru oricare două elemente $a, b \in E$ are loc $r_1(a) < r_1(b)$ numai dacă $r_2(a) < r_2(b)$;
- pentru oricare două elemente $a, b \in E$ din egalitatea $r_1(a) = r_1(b)$ rezultă $r_2(a) = r_2(b)$.

Se constată că un clasament care este o rafinare a altui clasament are mai multe locuri distincte decât acesta.

TEOREMA 1.4.2. Clasamentul asociat relației de preordine totală R_1 este o rafinare a clasamentului asociat relației de preordine totală R_2 dacă și numai dacă partiția ordonată corespunzătoare relației R_1 este o rafinare a partiției ordonate corespunzătoare relației R_2 .

Exemplul 1.4.2.

Să reluăm exemplul 1.3.1 și să presupunem că la o altă administrare a testului numărul erorilor comise de elevi este cel prezentat în tabelul următor:

Elev	e_1	e_2	e_3	e_4	e_5	e_6	e_7
Nr. erori	5	7	4	6	2	10	9

Relația de preordine totală R_1 pe mulțimea elevilor fiind definită la fel ca în exemplul 1.3.1, partiția ordonată corespunzătoare ei este: $(\{e_5\}, \{e_3\}, \{e_1\}, \{e_4\}, \{e_2\}, \{e_7\}, \{e_6\})$.

În acest caz, clasamentul elevilor este următorul:

Elev	Loc
e_5	1
e_3	2
e_1	3
e_4	4
e_2	5
e_7	6
e_6	7

Acest clasament este o rafinare a clasamentului corespunzător relației de preordine totală R din exemplul I.3.1.

I.5. COMPARAREA A DOUĂ PARTIȚII

Fie E o mulțime și
 $\{A_1, A_2, \dots, A_n\}$

și

$\{B_1, B_2, \dots, B_m\}$
două partiții ale mulțimii E .

Pentru a vedea cât de asemănătoare sunt cele două partiții, se construiește un tablou de frecvențe de următoarea formă:

Prima partiție	A doua partiție $B_1 \dots B_j \dots B_m$	
A_1	$f_{11} \dots f_{1j} \dots f_{1m}$	$f_{1.}$
...
A_i	$f_{i1} \dots f_{ij} \dots f_{im}$	$f_{i.}$
...
A_n	$f_{n1} \dots f_{nj} \dots f_{nm}$	$f_{n.}$
	$f_{.1} \dots f_{.j} \dots f_{.m}$	$f = N$

unde f_{ij} reprezintă numărul de obiecte care aparțin clasei A_i din prima partiție și clasei B_j din cea de-a doua partiție, pentru $i=1,2,\dots,n$ și $j=1,2,\dots,m$.

Cele două partiții concordă perfect dacă sunt identice, adică dacă ele conțin același număr de clase ($m=n$) și fiecare clasă dintr-o partiție este egală cu

o clasă din cealaltă partiție. Aceasta înseamnă că în tabelul de frecvențe de mai sus pe fiecare linie și pe fiecare coloană există câte o singură frecvență nenulă.

O posibilitate de a cerceta dacă cele două partiții au legătură între ele o oferă testul χ^2 .

Se calculează valoarea variabilei χ^2 cu ajutorul formulei:

$$\chi^2 = \sum_{i=1}^n \sum_{j=1}^m \frac{(f_{ij} - t_{ij})^2}{t_{ij}} \quad (I.5.1)$$

unde, pentru $i=1,2,\dots,n$ și $j=1,2,\dots,m$:

f_{ij} sunt frecvențele observate, citite din tabel;

t_{ij} reprezintă frecvențele teoretice, calculate cu formula:

$$t_{ij} = \frac{f_{i.} f_{.j}}{f} \quad (I.5.2)$$

Se compară valoarea χ^2 astfel calculată cu valoarea $\chi^2_{(n-1)(m-1), \alpha}$ citită din tabelul legii χ^2 pentru pragul α fixat și $(n-1)(m-1)$ grade de libertate.

Dacă $\chi^2 > \chi^2_{(n-1)(m-1), \alpha}$ atunci se respinge ipoteza că cele două partiții sunt independente probabilistic între ele și se acceptă ipoteza că între ele există legătură.

O altă metodă de comparare a partițiilor este cea propusă de Huberty (1984), aplicabilă în cazul când $m=n$ și fiecare clasă B_i corespunde clasei A_i , $i=1,2,\dots,n$. Procedeuul său este util atunci când se consideră că partiția $\{A_1, A_2, \dots, A_n\}$ este corectă și se dorește să se aprecieze calitatea celeilalte partiții prin comparare cu aceasta. De exemplu, când se construiește un test psihologic pentru determinarea tipului temperamental al subiecților, pentru a aprecia dacă testul funcționează corect, se vor compara rezultatele furnizate de el cu cele date de un alt test care este apreciat ca fiind bun.

Metoda propusă de Huberty se bazează pe compararea frecvenței relative a obiectelor plasate în aceeași clasă în ambele partiții:

$$RS = \frac{\sum_{i=1}^n f_{i,i}}{f} \quad (I.5.3)$$

numită "raportul succesului", cu valoarea S care se consideră că s-ar obține pentru acest raport dacă prima partiție este corectă, iar a doua ar fi realizată la întâmplare.

Modul de calcul al valorii S depinde de ipotezele care se fac.

A. Se presupune că rapoartele $f_i/f_{..}$, $i=1,2,\dots,n$, reflectă frecvențele relative reale, în populația din care provine mulțimea de obiecte clasificate, ale situațiilor cuprinse în clasele A_i , $i=1,2,\dots,n$.

În acest caz probabilitatea ca un obiect luat la întâmplare din populație să aparțină clasei A_i poate fi considerată egală cu $f_i/f_{..}$, iar frecvența absolută a obiectelor din clasa A_i care vor fi repartizate în clasa B_i datorită hazardului va fi $f_i f_i/f_{..}$, $i=1,2,\dots,n$. Pentru întreaga mulțime de obiecte clasificate frecvența relativă a concordanțelor datorate norocului va fi

$$S = \frac{\sum_{i=1}^n f_i^2}{f_{..}^2} \quad (1.5.4)$$

Dacă $RS \leq S$, atunci se poate aprecia că asemănările celor două partiții sunt datorate întâmplării.

În caz contrar, se verifică ipoteza nulă ($RS < S$) aplicând un test statistic de comparare a unei frecvențe relative (RS) cu o valoare dată (S).

B. Dacă frecvențele relative $f_i/f_{..}$, $i=1,2,\dots,n$, diferă de cele întâlnite în realitate și una dintre ele este mult mai mare decât celelalte, atunci se calculează valoarea S pornind de la observația că s-ar putea obține un număr mare de concordanțe repartizând toate obiectele în clasa A_i pentru care $f_i/f_{..}$ are cea mai mare valoare. Deci se ia

$$S = \max \{f_i/f_{..} | i=1,2,\dots,n\} \quad (1.5.5)$$

și se compară frecvența relativă RS cu S procedând ca în cazul precedent.

O altă metodă de comparare a două partiții, aplicabilă în aceleași condiții ca și metoda lui Huberty, utilizează coeficientul K al lui Cohen (Bakeman, Gottman, 1986):

$$K = \frac{RS - S}{1 - S} \quad (1.5.6)$$

care compară raportul succesului RS , dat de formula 1.5.3, cu valoarea

$$S = \sum_{i=1}^n \frac{f_i f_i}{f_{..}^2} \quad (1.5.7)$$

Acest coeficient ia valori între -1 și +1, concordanța perfectă a celor două partiții fiind indicată de valoarea maximă, +1.

1.6. COMPARAREA A DOUĂ CLASAMENTE

Compararea a două clasamente se bazează pe compararea partițiilor ordonate corespunzătoare relațiilor de preordine totală cărora le sunt asociate clasamentele.

Fie E o mulțime cu n elemente, R_1 și R_2 două relații de preordine totală pe E și

$$(A_1, A_2, \dots, A_p) \quad (1.6.1)$$

$$\text{și} \quad (B_1, B_2, \dots, B_m) \quad (1.6.2)$$

partițiile ordonate corespunzătoare lor.

Pentru a compara clasamentele asociate relațiilor R_1 și R_2 se compară partițiile ordonate (1.6.1) și (1.6.2).

În acest scop, se poate utiliza metoda lui Huberty și coeficientul K al lui Cohen.

În plus, în lucrările de statistică sunt definiți mai mulți coeficienți care permit compararea a două partiții ordonate, și care, având valori cuprinse între -1 și 1 (sau între 0 și 1) sunt ușor de interpretat.

Pentru fiecare element $e \in E$, notăm cu $r_1(e)$, respectiv cu $r_2(e)$, rangul elementului e în cele două partiții ordonate, (1.6.1) și (1.6.2).

Unul dintre coeficienții care servesc la compararea a două clasamente, cunoscut sub numele de **coeficientul de corelație a rangurilor al lui Spearman** este definit prin formula:

$$\rho = \frac{X_1 + X_2 - D}{2\sqrt{X_1 X_2}} \quad (1.6.3)$$

unde:

$$D = \sum_{e \in E} \{r_1(e) - r_2(e)\}^2 \quad (I.6.4)$$

$$X_i = \frac{n^3 - n}{12} - \frac{1}{12} \sum_{e \in E} \frac{1}{n_i(e)} \{n_i^3(e) - n_i(e)\} \quad (I.6.5)$$

$n_i(e)$ reprezintă numărul de elemente care au același rang ca și e în partiția ordonată corespunzătoare relației de preordine totală R_i , pentru $i=1,2$.

Dacă R_1 și R_2 sunt relații de ordine totală, atunci fiecare clasă de echivalență din cele două partiții ordonate conține un singur element, prin urmare $n_1(e)=n_2(e)=1$ pentru orice element $e \in E$. Deci:

$$\rho = 1 - \frac{6D}{n^3 - n} \quad (I.6.6)$$

Coeficientul ρ ia valoarea maximă, egală cu 1, dacă și numai dacă partițiile ordonate (I.6.1) și (I.6.2) coincid. Dacă R_1 și R_2 realizează clasamente "inverse" în sensul că $r_1(e)=n+1-r_2(e)$ pentru orice element $e \in E$, atunci $\rho=-1$.

Atunci când fiecare dintre partițiile ordonate (I.6.1) și (I.6.2) este asociată unei variabile măsurată pe scală de interval sau de raport, iar fiecare clasă de echivalență corespunde unei valori (măsurii) a acestei variabile, pentru compararea celor două partiții ordonate se poate utiliza coeficientul de corelație liniară. În acest scop se întocmește un tabel de frecvențe de următoarea formă:

Variabila X	Variabila Y	
	$y_1 \dots y_i \dots y_m$	
x_1	$f_{1,1} \dots f_{1,j} \dots f_{1,m}$	$f_{1.}$
...
x_i	$f_{i,1} \dots f_{i,j} \dots f_{i,m}$	$f_{i.}$
...
x_p	$f_{p,1} \dots f_{p,j} \dots f_{p,m}$	$f_{p.}$
	$f_{.1} \dots f_{.j} \dots f_{.m}$	$f = N$

unde f_{ij} reprezintă numărul de obiecte care au valoarea x_i pentru variabila X (aparțin clasei A_i din prima partiție) și valoarea y_j la variabila Y (aparțin clasei B_j din cea de-a doua partiție), pentru $i=1,2,\dots,p$ și $j=1,2,\dots,m$.

Coeficientul de corelație liniară r între variabilele X și Y se calculează cu formula:

$$r = \frac{\sum_{i=1}^p \sum_{j=1}^m x_i y_j f_{ij} - N m(X) m(Y)}{N \sigma(X) \sigma(Y)} \quad (I.6.7)$$

unde:

$m(X)$ și $m(Y)$ sunt mediile variabilelor X și Y :

$$m(X) = \frac{\sum_{i=1}^p x_i f_{i.}}{N} \quad (I.6.8)$$

$$m(Y) = \frac{\sum_{j=1}^m y_j f_{.j}}{N} \quad (I.6.9)$$

$\sigma(X)$ și $\sigma(Y)$ sunt abaterile standard ale variabilelor X și Y :

$$\sigma(X) = \sqrt{\frac{\sum_{i=1}^p (x_i - m(X))^2 f_{i.}}{N-1}} \quad (I.6.10)$$

$$\sigma(Y) = \sqrt{\frac{\sum_{j=1}^m (y_j - m(Y))^2 f_{.j}}{N-1}} \quad (I.6.11)$$

Acest coeficient arată în ce măsură norul de puncte (x_i, y_j) , $i=1,2,\dots,p$ și $j=1,2,\dots,m$ se grupează în jurul unei drepte.

Când norul de puncte este format din puncte așezate de-a lungul unei drepte, coeficientul de corelație liniară are valoarea +1 (dacă dreapta are pantă pozitivă) sau -1 (dacă dreapta are pantă negativă). Cu cât acest nor este mai

împrăștiat, cu atât valoarea coeficientului de corelație liniară este mai apropiată de 0.

Cu ajutorul testului Student (t) se poate verifica dacă valoarea unui coeficient de corelație liniară este semnificativ diferită de 0. Pentru aceasta se calculează valoarea t , cu ajutorul formulei:

$$t = r \sqrt{\frac{N-2}{1-r^2}} \quad (1.6.12)$$

unde:

N este volumul mulțimii E ;

r este valoarea calculată a coeficientului de corelație liniară.

Din tabelul legii Student se citește valoarea $t_{N-2, \alpha}$ corespunzătoare pragului de probabilitate α fixat și la $N-2$ grade de libertate. Dacă $t > t_{N-2, \alpha}$ atunci se respinge ipoteza că valoarea coeficientului r este nesemnificativă statistic (nu diferă semnificativ de 0).

1.7. AGREGAREA CLASAMENTELOR

Prin agregarea mai multor clasamente într-unul singur se înțelege definirea unei ierarhii a elementelor unei mulțimi E plecând de la o mulțime dată de clasamente inițiale.

Fie E o mulțime finită, m un număr natural, $m \geq 2$, și S_E mulțimea relațiilor de preordine totală existente pe E .

Notăm cu

$$S_E^m = \underbrace{S_E \times S_E \times \dots \times S_E}_{\text{de } m \text{ ori}}$$

mulțimea m -uplurilor de preordini totale ale elementelor din E .

DEFINIȚIA 1.7.1. Se numește **procedeu de agregare** o funcție $g: S_E^m \rightarrow S_E$.

Adică, un procedeu de agregare a clasamentelor construiește un clasament pe mulțimea E cu ajutorul a m clasamente pe E .

Trebuie reținut că nu orice funcție care construiește o relație binară completă R pe mulțimea E pornind de la m preordini totale pe E este un

procedeu de agregare, întrucât nu se asigură totdeauna tranzitivitatea relației R . Un exemplu de acest fel îl constituie metoda deciziei majoritare (Arrow, 1963), care definește o relație R în mulțimea E utilizând m relații de preordine totală pe E , R_1, R_2, \dots, R_m astfel:

$$\forall a, b \in E : aRb \Leftrightarrow \sum_{i=1, m} 1_{aR_i b} \geq \sum_{i=1, m} 1_{bR_i a} \quad (1.7.1)$$

Adică a este preferat lui b în clasamentul final (are loc relația aRb) dacă numărul ierarhiilor în care a este preferat lui b nu este mai mic decât numărul ierarhiilor în care b este preferat lui a .

Este binecunoscut următorul exemplu, construit de marchizul de Condorcet în 1785 (Păun, 1987).

Se consideră trei candidați, a , b și c și 60 de alegători care propun, fiecare, câte o relație de ordine totală pe mulțimea $\{a, b, c\}$ și, deci, o ierarhie a elementelor acestei mulțimi în care nu pot să apară doi candidați pe același loc. Să presupunem că s-au obținut următoarele ierarhii:

Locul	Frecvența
1 2 3	
a b c	de 23 de ori
b c a	de 17 ori
b a c	de 2 ori;
c b a	de 8 ori;
c a b	de 10 ori.

Să aplicăm metoda deciziei majoritare, folosind simbolul $>$ pentru a arăta că un element se află înaintea altuia în ierarhie și notând cu R relația binară construită:

$$\begin{aligned} a > b \text{ de } 33 \text{ de ori și } b > a \text{ de } 27 \text{ de ori} &\Rightarrow aRb \\ b > c \text{ de } 42 \text{ de ori și } c > b \text{ de } 18 \text{ de ori} &\Rightarrow bRc \\ a > c \text{ de } 25 \text{ de ori și } c > a \text{ de } 35 \text{ de ori} &\Rightarrow cRa \end{aligned}$$

Deci relația R nu este tranzitivă.

Două procedee de agregare, definite pentru cazul când ierarhiile de pornire corespund unor relații de ordine totală sunt metoda lui Borda (Păun, 1987) și metoda lui Onicescu (1970).

Metoda lui Borda acordă $m, m-1, \dots, 2, 1$ puncte locurilor $1, 2, \dots, m$ din ierarhiile de plecare, însumează punctele obținute de fiecare element și ordonează descrescător punctajele totale.

Metoda lui Onicescu este similară metodei anterioare, cu deosebirea că locurilor $1, 2, \dots, m$ le acordă un număr de puncte:

$$1, \frac{1}{2}, \frac{1}{2^2}, \dots, \frac{1}{2^{m-1}}$$

I.8. OPERAȚIILE CLASIFICĂRII

Clasificarea este o operație logică care acționează asupra unei mulțimi de obiecte (sau ființe sau fenomene), fie pentru a le grupa, formând **clase**, fie pentru a le repartiza în clase existente.

Gruparea obiectelor în clase, prin clasificare, se face astfel încât gradul de asemănare (de similaritate) să fie mare între membrii aceleiași clase și mic între membrii unor clase diferite.

În foarte multe cazuri clasele sunt disjuncte două câte două, determinând o *partiție* a mulțimii obiectelor clasificate.

Alteori, prin clasificare se definește o *partiție ordonată* și o *ierarhie*. Această operație se numește **ierarhizare**. Clasele formate în acest caz sunt alcătuite din elemente echivalente între ele.

În timpul constituirii claselor se stabilește și o descriere a lor, sub forma prototipurilor. Un *prototip* este un obiect ideal al clasei (de exemplu, poate fi centrul de greutate al obiectelor grupate în clasă) sau o mulțime de obiecte ideale care îndeplinesc o anumită condiție (de exemplu, valorile caracteristicilor lor verifică o ecuație). Termenul de "obiect ideal" desemnează un element care nu este necesar să aparțină mulțimii obiectelor clasificate, dar care este descris în același mod ca și acestea, adică prin aceleași caracteristici.

Operația de determinare a clasei în care este cel mai potrivit să fie repartizat un obiect care nu aparține mulțimii de obiecte pe baza cărora s-au format clasele este denumită **clasare** sau **identificare**. Ea acționează de obicei prin compararea obiectului cu prototipurile claselor și alegerea clasei al cărei prototip seamănă cel mai mult cu obiectul în cauză.

ANEXA II CÂTEVA NOȚIUNI MATEMATICE ȘI STATISTICE CARE INTERVIN ÎN TEORIA TESTELOR PSIHOLOGICE

II.1 DEFINIȚIA VARIABILEI

O **variabilă** este o clasă generală sau o categorie de obiecte, evenimente, proprietăți sau situații. Câteva exemple de variabile sunt: performanța la un test de atenție, inteligența, sexul, intensitatea zgomotului. În cadrul categoriei se pot observa diverse moduri de existență sau de manifestare a variabilei. De exemplu, pentru variabila "situația școlară la sfârșitul anului școlar, pentru elevi" sunt posibile cazurile "promovat", "corigent" și "repetent". Acestea se numesc **niveluri** sau **valori** ale variabilei.

O variabilă care are numai două valori se numește **variabilă dihoto-mică** sau **variabilă booleană** sau **variabilă binară**. Așa sunt, de exemplu, variabila "sex" și variabila asociată unui item la care răspunsurile posibile sunt "da" și "nu".

Unele variabile, precum sexul sau culoarea ochilor, au valori **calitative** (exprimate prin cuvinte), altele, cum sunt vârsta și înălțimea, au valori **cantitative** (exprimate prin numere).

În cazul variabilelor cu valori numerice sunt posibile două situații:

(1) Variabila poate lua numai anumite valori numerice. Între oricare două valori ale variabilei există doar un număr fixat de valori posibile. Mulțimea tuturor valorilor variabilei este finită sau numărabilă. În acest caz se spune că variabila este **discretă**. De exemplu, numărul de copii al unei familii sau numărul de persoane dintr-o instituție sunt variabile discrete.

(2) Variabila poate lua orice valoare numerică dintr-un interval finit sau infinit dat. Între oricare două valori ale variabilei există o infinitate de valori posibile. În acest caz se spune că variabila este **continuă**.

II.2. VARIABILE DEPENDENTE, VARIABILE INDEPENDENTE

Variabila **dependentă** este variabila ale cărei valori depind de o altă variabilă, numită variabilă **independentă**. La rândul ei, variabila independentă

poate să depindă de o altă variabilă, fiind, deci, variabilă dependentă. Relația de dependență se referă strict la două variabile.

Există două tipuri de variabile independente:

a. **variabilele manipulate** sunt cele pe care experimentatorul le controlează prin manipulare activă pentru a urmări schimbările măsurabile apărute la variabila dependentă;

b. **variabilele nemanipulate** sunt variabile de clasificare: în funcție de valorile lor se grupează unitățile statistice și apoi se verifică dacă între grupele formate există deosebiri în privința valorilor variabilei dependente.

Exemple.

1. În cercetările asupra învățării, variabila independentă manipulată poate fi intervalul de timp în care este prezent un stimul, iar variabila dependentă este precizia cu care subiecții recunosc stimulul. O altă variabilă independentă manipulată poate fi complexitatea stimulului, iar variabila dependentă, timpul necesar subiecților pentru recunoașterea stimulului.

2. Se dorește să se verifice ipoteza că deprinderile de rezolvare a problemelor morale depind de apartenența religioasă. În acest scop se formează grupe de persoane, în funcție de religie, cărora li se administrează același test referitor la probleme morale. Se compară apoi rezultatele între grupe. În acest caz apartenența religioasă este variabila independentă nemanipulată.

Unii autori înțeleg prin variabilă independentă o variabilă care cauzează modificări ale variabilei dependente. Variabila dependentă este efect al variabilei independente. Dar, dacă nu se controlează prin manipulare variabila independentă, este dificil de tras concluzii asupra existenței relației cauză-efect între variabila independentă și cea dependentă.

II.3. SCALE DE MĂSURĂ

Compararea a două persoane, X și Y , conduce adesea la formularea unor întrebări de genul: Este X mai frumoasă decât Y ? Cu cât este X mai înaltă decât Y ? De câte ori este câștigul lui Y mai mare decât cel al lui X ?

Descrierea unui obiect X poate fi făcută mai sugestivă raportând calitățile sale la cele ale populației de obiecte din care face parte. Aceasta generează întrebări precum: Obiectul X este cu mult mai scump decât celelalte? Obiectul X este mai greu sau mai ușor decât celelalte? ș.a.m.d.

Formularea răspunsurilor la întrebări ca cele de mai sus este mult simplificată dacă se dau valori numerice caracteristicilor analizate, astfel încât

compararea obiectelor în privința fiecărei caracteristici să se reducă la compararea unor numere. Se realizează, în acest fel, *măsurarea* persoanelor sau a obiectelor.

Există mai multe definiții ale măsurării. Majoritatea acestora sunt de forma următoare: "**măsurarea** constă în atribuirea de x lui y în conformitate cu z ". În funcție de autor:

- | | |
|-----|--|
| x | reprezintă "numere", "valori", "scoruri" sau "simboluri abstracte"; |
| y | este înlocuit cu "obiecte", "evenimente", "lucruri", "situații", "indivizi", "comportamente", "observații", "attribute", "proprietăți" sau "răspunsuri"; |
| z | primește semnificația de "regulă specifică" sau "regulă de reprezentare" (Michell, 1997). |

În cadrul acestei lucrări, prin măsurare a unei caracteristici la o mulțime de obiecte înțelegem atribuirea de numere obiectelor studiate, conform unei reguli bine precizate, prin care se urmărește ca anumite relații existente între obiecte, în privința caracteristicii respective, să existe și între numerele atribuite lor. Mai precis, dacă în mulțimea obiectelor este definită o anumită relație (de exemplu, o relație de ordine), atunci trebuie să existe o relație similară în mulțimea măsurilor elementelor, iar dacă două obiecte sunt în relația considerată atunci măsurile lor trebuie să fie în relația similară ei. De pildă, dacă din punctul de vedere al caracteristicii cercetate un obiect X este superior unui obiect Y , atunci măsura obiectului X trebuie să fie mai mare decât măsura obiectului Y .

Înseamnă, deci, că măsurarea stabilește un izomorfism între mulțimea obiectelor și mulțimea măsurilor obiectelor. În asemenea condiții se pot efectua diverse operații asupra măsurilor obiectelor, iar concluziile desprinse se extind asupra obiectelor.

Din punct de vedere matematic, măsurarea este o funcție care atașează fiecărui element din mulțimea de obiecte (mulțimea de definiție a funcției), conform unor anumite reguli, un număr și numai unul (din mulțimea în care funcția ia valori), numit "măsura elementului".

Fiecare sistem de reguli impus definește câte un tip de măsurare sau câte un tip de **scală**. În general sunt utilizate patru tipuri de scale de măsură. În ordinea crescătoare a preciziei lor, acestea sunt:

- scala nominală;
- scala ordinală;
- scala de interval;
- scala de raport.

Regulile impuse pentru definirea unui tip de scală le includ, totdeauna, pe cele impuse pentru definirea scalelor aflate pe un nivel inferior. De exemplu,

dacă s-a putut defini o scală de interval, atunci s-ar fi putut defini și o scală ordinală sau una nominală.

Trebuie reținute două idei:

(1) Alegerea unui anumit nivel de măsurare pentru o mulțime de obiecte este dependentă de numărul și tipul relațiilor existente între obiecte și care prezintă interes pentru cercetător. Evident, aceste relații se referă la caracteristica analizată.

(2) Prelucrările statistice care pot fi efectuate asupra măsurilor obiectelor pentru a cunoaște mai bine obiectele sunt dependente de tipul de scală utilizat. Cu cât scala de măsură este mai precisă, cu atât prelucrările statistice permise sunt mai complexe și conduc la concluzii mai numeroase.

Notăm în continuare cu A mulțimea obiectelor (sau a indivizilor sau a fenomenelor), cu B mulțimea de numere din care se dau valori măsurilor elementelor din A și cu $m: A \rightarrow B$ funcția care realizează măsurarea. Deci, pentru orice obiect $a \in A$, $m(a)$ reprezintă măsura sa.

Măsurarea pe o scală nominală poate fi realizată ori de câte ori caracteristica studiată permite împărțirea elementelor mulțimii A în clase disjuncte două câte două, adică atunci când există o partiție a mulțimii A . În fiecare clasă sunt grupate toate obiectele care pot fi considerate echivalente în privința acestei caracteristici.

De pildă, variabilele referitoare la categoria socio-profesională sau la sexul persoanelor pot fi măsurate pe scală nominală.

Regula impusă la măsurarea nominală este ca toate elementele dintr-o aceeași clasă de echivalență să aibă aceeași măsură și o valoare a măsurii să apară numai la elemente echivalente între ele. Adică:

R1. Două elemente, $a, b \in A$, au măsuri egale ($m(a) = m(b)$) dacă și numai dacă ele sunt echivalente ($a = b$).

O posibilitate de definire a unei scale nominale constă în atribuirea de numere distincte claselor de obiecte. Toate obiectele unei clase vor avea aceeași măsură, egală cu numărul atașat clasei. De exemplu, prin "numărarea" elementelor din A se realizează o măsurare nominală; în acest caz fiecare clasă este constituită dintr-un singur obiect. Numerele pe care le primesc jucătorii unei echipe de fotbal în cadrul unui joc sunt de asemenea măsuri nominale.

Măsura unui element este o "etichetă" atribuită elementului, care servește doar la denumirea, identificarea sau clasificarea acestuia. Ea este numerică, dar ar fi putut fi exprimată și prin cuvinte. Din acest motiv nu are sens să se efectueze calcule aritmetice cu măsurile nominale.

Prin măsurare nominală se obține o singură informație asupra elementelor mulțimii A : două elemente care au aceeași măsură sunt echivalente între ele. Se pot schimba între ele numerele atribuite claselor fără ca informația pe care ne-o furnizează o măsurare nominală să se schimbe.

Măsurarea pe o scală ordinală impune ca elementele mulțimii A să poată fi aranjate în ordine, dar cu permisiunea ca mai multe obiecte să ocupe același loc. De exemplu, muncitorii unei secții pot fi ordonați ținând cont de randamentul lor, de la cel mai eficient la cel mai puțin eficient.

Convenim să spunem că un element $a \in A$ este "superior" unui element $b \in A$ dacă în șirul ordonat format elementul a se află înaintea lui b . Notăm acest lucru prin $a > b$. Relația $>$ în cazul unei caracteristici concrete se poate traduce prin "este mai mare decât" sau "este mai bun decât" sau "este preferat lui" etc.

Dacă două elemente, $a, b \in A$, ocupă același loc în șirul ordonat, vom spune că a și b sunt "echivalente" și vom nota $a = b$. Relația $=$ poate însemna, în funcție de context, "este la fel de mare ca" sau "este la fel de bun ca" sau "este apreciat la fel ca" etc.

Prin urmare, măsurarea pe scală ordinală poate fi realizată dacă în mulțimea A există o relație de preordine totală, pe care o notăm cu \geq . Oricare două elemente, a și b din A , îndeplinesc una și numai una dintre următoarele condiții:

$a > b$, dacă $a \geq b$ și $\text{non}(b \geq a)$;

$b > a$, dacă $b \geq a$ și $\text{non}(a \geq b)$;

$a = b$, dacă $a \geq b$ și $b \geq a$.

Măsurării ordinale i se impun **regulile** R1, de la scala nominală, și:

R2. Pentru oricare două elemente $a, b \in A$ aflate în relația $a > b$ măsura elementului a este mai mare decât măsura elementului b : $m(a) > m(b)$.

Notăm cu (A_1, A_2, \dots, A_k) partiția ordonată corespunzătoare preordinii totale \geq . Aici, pentru $i = 1, 2, \dots, k$, fiecare mulțime A_i conține elemente aflate între ele în relația $=$ și toate elementele din mulțimea A_i sunt în relația $>$ cu toate elementele mulțimilor $A_{i+1}, A_{i+2}, \dots, A_k$.

Un algoritm simplu pentru definirea unei scale ordinale este următorul:

(1) se aleg k numere, $m_1 > m_2 > \dots > m_k$;

(2) tuturor elementelor din clasa A_i li se atribuie numărul m_i ;

$m(a) = m_i$, pentru orice $a \in A_i$, $i = 1, 2, \dots, k$.

Se realizează de asemenea o măsurare pe scală ordinală dacă fiecărui element a din A i se atașează numărul $N - r(a) + 1$, unde N este volumul mulțimii A , iar $r(a)$ este rangul elementului a în relația de preordine totală \geq .

Exemplu.

Să presupunem că mulțimea $A = \{a, b, c, d, e, f, g\}$ este formată din membrii unei echipe de handbal, care au primit după un meci următoarele calificative:

a : slab; b : foarte bun; c : slab;
 d : foarte slab; e : mediocru; f : slab;
 g : slab.

Pe baza calificativelor, în mulțimea jucătorilor se poate defini o relație de preordine totală, căreia îi corespunde partiția ordonată $(\{b\}, \{e\}, \{a, c, f, g\}, \{d\})$. Se determină rangul fiecărui jucător și măsurile calculate cu ajutorul rangului:

Jucător	Calificativ	Rang	Măsura
b	foarte bun	1	7
e	mediocru	2	6
a	slab	4.5	3.5
c	slab	4.5	3.5
f	slab	4.5	3.5
g	slab	4.5	3.5
d	foarte slab	7	1

Măsurarea ordinală dă următoarele informații:

- două elemente care au aceeași măsură sunt echivalente (nici una nu îi este superioară celeilalte);
- dintre două elemente cu măsuri neegale, cel care are măsura mai mare este superior celui alt.

Trebuie reținut că măsurarea ordinală nu spune nimic despre "distanța" existentă între două elemente ale mulțimii A . Dacă $m(a) - m(b) = m(c) - m(d)$ nu înseamnă că a este tot cu atâta superior lui b cu cât c este superior lui d .

Transformând măsurile obiectelor din mulțimea A printr-o funcție crescătoare, se obține o altă măsurare ordinală a elementelor din A , întrucât noile măsuri păstrează ordinea obiectelor.

Măsurarea pe o *scală de interval* poate fi utilizată atunci când în mulțimea A există un clasament și, în plus, se poate aprecia "distanța" dintre oricare două elemente în privința caracteristicii studiate. Această "distanță" este o funcție

$$d : A \times A \rightarrow \mathbb{R}$$

care atribuie fiecărei perechi de elemente din A , (a, b) , un număr nenegativ, $d(a, b) \geq 0$, și îndeplinește următoarele trei condiții:

C1. "Distanța" dintre oricare două elemente echivalente este egală cu zero: $a=b \Rightarrow d(a, b)=0$.

C2. Pentru oricare două elemente $a, b \in A$, "distanța" dintre a și b este egală cu "distanța" dintre b și a : $d(a, b)=d(b, a)$.

C3. Pentru oricare trei elemente $a, b, c \in A$, între care există relațiile ($a > b$ sau $a=b$) și ($b > c$ sau $b=c$), are loc egalitatea:
 $d(a, c) = d(a, b) + d(b, c)$

Unei scale de interval i se cere ca, pe lângă regulile R1 și R2 impuse scalei ordonate, să îndeplinească și următoarea **regulă**:

R3. Pentru oricare patru elemente $a, b, c, e \in A$ are loc
 $|m(a) - m(b)| = |m(c) - m(e)|$ dacă și numai dacă $d(a, b) = d(c, e)$.

Această regulă înseamnă că scala de interval atribuie unor obiecte egal distanțate între ele în privința caracteristicii cercetate, măsuri echidistante.

Un exemplu de măsurare pe scală de interval îl constituie evaluarea inteligenței unor personalități, realizată de C. Cox pe baza parcurgerii unui material bibliografic vast. Au fost atribuite următoarele măsuri: 145 pentru Napoleon, 155 pentru Rembrandt, 165 pentru Beethoven și 185 pentru Galileo Galilei. Prin acești coeficienți de inteligență, Cox a vrut să afirme că "diferența dintre Rembrandt și Napoleon este egală cu cea dintre Beethoven și Rembrandt", iar "Galileo Galilei a fost mai inteligent decât Beethoven cu atâta cu cât Beethoven a fost mai inteligent decât Napoleon".

În cazul scalelor de interval, faptul că un element are măsură egală cu zero nu înseamnă că el este lipsit de caracteristica cercetată.

Exemplul tipic de scală de interval se întâlnește la măsurarea temperaturii. După cum se știe, temperatura este un parametru ce caracterizează starea de încălzire a unui sistem fizic și se determină pe baza variației unor mărimi fizice (volum, rezistență electrică, forță electromotoare etc.), astfel încât unor variații egale ale mărimii fizice le corespund variații egale de temperatură. Termometrul cu care se înregistrează valorile temperaturii este gradat în unități echidistante. Indiferent ce reprezintă distanța dintre două gradații succesive (un grad Celsius sau un grad Fahrenheit sau altceva) diferența de temperatură corespunzătoare distanței dintre două gradații succesive este la fel de mare, fie că este vorba despre gradațiile 2 și 3 sau 10 și 11. Se obișnuiește să se atribuie unei temperaturi măsură zero, dar aceasta nu înseamnă absența totală a temperaturii. Zero grade Celsius corespunde, de exemplu, la 32 de grade Fahrenheit. Atât scala Celsius, cât și scala Fahrenheit realizează măsurări de interval.

Un alt exemplu de măsurare pe scală de interval îl reprezintă calendarele. Unitatea de măsură pentru timp (an, lună, zi) este constantă în fiecare calendar. Dar, în calendarul evreiesc numărarea anilor se face începând cu "crearea lumii", a cărei dată este estimată prin secolul al IV-lea î.C. Anului 1995 îi corespunde în calendarul evreiesc anul 5755.

Din aceste exemple se observă că măsurarea de tip interval fixează arbitrar punctul zero (mărimea caracteristicii căreia să i se atribuie măsura zero) și lungimea unității de măsură ("distanța" dintre două niveluri ale caracteristicii căreia să îi corespundă o diferență a măsurilor egală cu unu).

În cazul măsurării pe scală de interval, dacă măsurile obiectelor sunt transformate printr-o funcție liniară:

$$m_{nou}(a) = p + qm(a), \text{ pentru orice element } a \in A$$

unde:

p și q sunt numere reale, $q > 0$,

$m(a)$ este măsura elementului a în scala de interval considerată,

$m_{nou}(a)$ este măsura nouă a elementului a

se obține de asemenea o măsurare de interval.

Această proprietate face posibilă compararea între ele a mai multor caracteristici ale unui obiect, măsurate pe scală de interval, chiar dacă fiecare dintre măsurători furnizează valori din alt interval.

Exemplu.

Se administrează unei clase de elevi teste de cunoștințe la gramatică, la matematică și la biologie, fiecare test furnizând scoruri într-un alt interval. De pildă, la gramatică se pot obține scoruri între 0 și 20, la matematică între 0 și 50, iar la biologie între 0 și 100. Pentru a aprecia, la care dintre cele trei discipline este mai bun fiecare elev este necesar să se efectueze transformări prin care scorurile testelor să fie aduse într-un același interval de valori, de exemplu să fie cuprinse între -10 și +10.

Măsurarea pe o *scală de raport* poate fi realizată ori de câte ori sunt îndeplinite condițiile pentru definirea unei scale de interval și, în plus, se poate aprecia în mod obiectiv când un element este lipsit de caracteristica cercetată.

De exemplu, se poate afirma cu siguranță că un vehicul care staționează este lipsit de viteză.

Scalei de raport i se impune ca, pe lângă regulile R1, R2 și R3 să îndeplinească și **regula** următoare:

R4. Un element $a \in A$ are măsura egală cu zero dacă și numai dacă el este lipsit de caracteristica cercetată.

Exemplu.

A. Moles a realizat un clasament pentru 250 de compozitori cunoscuți, în ordinea frecvenței cu care au fost interpretate creațiile lor în concerte, într-un interval de timp fixat. S-a constatat astfel că lucrările lui Schubert au fost incluse în 3.6% din programele de concert, cele ale lui Johann Strauss în 1.2%, iar cele ale lui Gershwin, în 0.3%. Deci, compozițiile lui Schubert sunt interpretate de 3 ori mai frecvent decât cele ale lui Johann Strauss și de 12 ori mai frecvent decât cele ale lui Gershwin. Variabila a cărei valoare pentru fiecare compozitor este frecvența cu care sunt interpretate creațiile sale, exprimată în procente sau în număr de cazuri, este măsurată pe scală de raport. Ea are valoarea 0 doar dacă nici una dintre compozițiile persoanei nu a fost inclusă în programele de concert.

Atunci când caracteristica măsurată reprezintă spațiu (lungime, suprafață, volum), masă sau timp trecut de la un eveniment fixat (de exemplu, durată de viață, timp de reacție) se poate stabili în mod obiectiv ce înseamnă "absența caracteristicii". Pentru asemenea variabile se pot defini măsurări de raport.

Deosebirea dintre o măsurare de interval și una de raport constă în faptul că, la măsurarea de interval, din egalitatea $m(a) = Km(b)$ nu se poate trage concluzia că a este "de K ori mai bun" decât b (în sensul relației $>$), dar acest lucru este posibil la măsurarea de raport. De pildă, indiferent dacă înălțimea persoanelor este măsurată în mm, m sau înch, un copil care are 160 cm este de două ori mai înalt decât unul care are 80 cm. Dacă o mașină se deplasează cu 60 km/oră iar alta cu 30 km/oră, se poate spune că prima merge de două ori mai repede decât a doua.

În cazul măsurării pe scală de raport, transformând măsurile obiectelor printr-o funcție liniară:

$$m_{nou}(a) = p + qm(a), \text{ pentru orice element } a \in A$$

cu $p, q \in \mathbb{R}$ și $q > 0$, dacă numărul p este diferit de zero nu se va mai obține o măsurare pe scală de raport, ci doar o măsurare pe scală de interval. De această observație trebuie să se țină seama atunci când se efectuează transformarea valorilor unei variabile în cote standard.

Gruparea valorilor unei variabile măsurate pe scală de interval sau de raport poate conduce la transformarea acesteia într-o variabilă măsurată pe scală ordinală.

Exemplu.

Variabila "vârstă", care are ca valoare, pentru fiecare persoană, numărul de ani împliniți, este măsurată pe scală de raport. Dacă se stabilesc trei grupe:

"sub 25 de ani", "între 25 și 60 de ani" și "peste 60 de ani", cărora li se atribuie valorile 1, 2 și 3, atunci variabila va fi măsurată pe scală ordinală.

Scala de interval și cea de raport sunt denumite **scale metrice**. Variabilele ale căror valori se obțin printr-o măsurare pe scală de interval sau pe scală de raport se numesc **variabile de tip scor**.

Alegerea unui tip de scală pentru măsurarea unei variabile trebuie să aibă în vedere, în primul rând, tipul de prelucrări statistice care ar trebui efectuate asupra variabilei. Dacă relațiile cunoscute între elementele mulțimii A nu permit măsurarea variabilei pe scala care ar fi necesară calculelor statistice respective, se va realiza o măsurare pe o scală inferioară dar cât mai apropiată de aceasta.

Atunci când "obiectele" care se măsoară sunt comportamente, apare o problemă suplimentară, cunoscută sub denumirea de "reactivitate". Se spune că o măsură este reactivă dacă operația de măsurare schimbă comportamentul subiecților (Cozby, 1993). Când persoanele se știu observate, comportamentul lor poate să difere de cel pe care l-ar avea în condiții normale. Chiar și parametrii fiziologici își pot schimba valorile din cauza prezenței echipamentelor de înregistrare a lor. Din acest motiv, se recomandă ca înainte de a efectua măsurătorile să i se lase timp subiectului să se obișnuiască cu prezența observatorului sau a aparatului de măsură.

II.4. DISTANȚE

Fie E o mulțime de vectori din spațiul euclidian M -dimensional ($M \geq 2$) și $x, y \in E$ două elemente oarecare:

$$x = (x_1, x_2, \dots, x_M)$$

$$y = (y_1, y_2, \dots, y_M)$$

De exemplu, E poate corespunde unei mulțimi de persoane care au răspuns la M teste psihologice. Pentru fiecare persoană există în E un vector, care are ca și componente scorurile realizate la teste, ordinea testelor fiind fixată (adică, cea de-a i -a componentă va conține scorul la testul i).

Pentru calculul distanței dintre x și y se pot folosi următoarele norme ale spațiului R^M :

- distanța Manhattan sau city block:

$$D_1(x, y) = \sum_{i=1}^M |x_i - y_i| \quad (II.4.1)$$

- distanța euclidiană:

$$D_2(x, y) = \sqrt{\sum_{i=1}^M (x_i - y_i)^2} \quad (II.4.2)$$

- distanța valorii maxime:

$$D_{\max}(x, y) = \max\{|x_i - y_i| : i=1, 2, \dots, M\} \quad (II.4.3)$$

- distanța Hamming, aplicabilă atunci când fiecare componentă a vectorilor are doar un număr finit de valori: distanța Hamming dintre vectorii x și y este dată de numărul componentelor în care cei doi vectori diferă.

- distanța lui Mahalanobis:

$$D_M(x, y) = \{(x-y)^T C^{-1} (x-y)\}^{1/2} \quad (II.4.4)$$

unde:

$(x-y)^T$ reprezintă transpusa vectorului coloană $x-y$;

$C = (c_{ij})_{i,j=1,2,\dots,M}$ este matricea de covarianță a celor M caracteristici pe mulțimea E .

Atunci când componentele vectorilor din mulțimea E sunt valorile a M variabile și aceste variabile sunt necorelate liniar două câte două pe mulțimea E , are loc $\sigma_{ij} = 0$ pentru $i, j = 1, 2, \dots, M, i \neq j$; deci matricea C^{-1} va fi o matrice diagonală, $C^{-1} = (c_{ij})_{i,j=1,2,\dots,M}$, cu $c_{ij} = 0$ și $c_{ii} = 1/\sigma_{ii}$, $i, j = 1, 2, \dots, M, i \neq j$.

Observații

1. În cazul distanței euclidiene și a distanței city block se poate obține aceeași valoare a distanței dintre doi vectori în două situații foarte deosebite: când diferă câte puțin valorile fiecărei componente și când diferă mult valorile unei singure componente, celelalte fiind egale. Din acest motiv ele nu sunt potrivite să fie folosite atunci când o diferență mare între valorile unei singure componente (indiferent care) trebuie să conducă la concluzia că obiectele se deosebesc mult între ele. În asemenea situații este mai potrivită distanța valorii maxime.

Exemplu.

Distanțele dintre vectorii:

$$x = (0, 0, 0, 0, 0, 0, 0, 0)$$

$$y = (1, 1, 1, 1, 1, 1, 1, 1)$$

$$z = (3, 0, 0, 0, 0, 0, 0, 0)$$

$$v = (9, 0, 0, 0, 0, 0, 0, 0)$$

sunt: $D_1(x, y) = D_1(x, v) = 9, D_1(x, z) = 3$

$$D_2(x, y) = D_2(x, z) = 3, D_2(x, v) = 9$$

$$D_{\max}(x, y) = 1, D_{\max}(x, z) = 3, D_{\max}(x, v) = 9$$

2. Distanța euclidiană, spre deosebire de distanța city block, "amplifică" distanțele dintre componente, făcând ca distanțele mari să pară și mai mari (pentru că în formula de calcul a distanței euclidiene diferențele valorilor componentelor se ridică la pătrat). Adică, pentru distanța euclidiană, o diferență mare între valorile unei componente contează mai mult decât diferențele mici sau medii între valorile a două componente.

Exemplu

Distanțele dintre vectorii

$$x = (0,0,0,0,0)$$

$$y = (1,1,1,1,0)$$

$$z = (1,1,2,0,0)$$

sunt:

$$D_1(x,y) = D_1(x,z) = 4$$

$$D_2(x,y) = 2 \quad D_2(x,z) = 2.449$$

3. Distanța Hamming este potrivită atunci când nu are importanță cât de mult se deosebesc între ele valorile componentelor, ci numai dacă ele sunt diferite.

Exemplu.

Distanțele Hamming dintre vectorii:

$$x = (0,0,0,0,0,0,0,0)$$

$$y = (1,1,1,1,1,1,1,1)$$

$$z = (3,0,0,0,0,0,0,0)$$

$$v = (9,0,0,0,0,0,0,0)$$

sunt:

$$D_H(x,y) = 9, \quad D_H(x,z) = D_H(x,v) = 1$$

4. Calculul distanței lui Mahalanobis ține cont și de felul în care valorile componentelor covariază (variază unele în raport cu altele), dar necesită cunoașterea tuturor vectorilor din mulțimea E și calculul covarianțelor componentelor.

II.5. COTE STANDARD, COTE STANDARDIZATE

Fie $X = \{x_1, x_2, \dots, x_n\}$ o mulțime de obiecte și V o variabilă care a fost măsurată, folosind o scală de interval sau de raport, la toate elementele mulțimii X .

Notăm:

m = media valorilor variabilei V pe mulțimea X :

$$m = \frac{\sum_{i=1}^n x_i}{n} \quad (II.5.1)$$

σ = abaterea standard a valorilor variabilei V pe mulțimea X :

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - m)^2}{n}} \quad (II.5.2)$$

Cota standard z corespunzătoare unei valori x_i din X se calculează cu formula:

$$z_i = \frac{x_i - m}{\sigma} \quad (II.5.3)$$

Cota standard z_i arată cu câte unități de abatere standard se distanțează valoarea x_i de media valorilor variabilei V , m .

Transformarea în cote standard nu modifică proporționalitatea intervalor scalei. Astfel, dacă x_i, x_j și x_k reprezintă trei valori ale variabilei V , iar z_i, z_j

$$\frac{x_i - x_j}{x_j - x_k} = \frac{z_i - z_j}{z_j - z_k} \quad (II.5.4)$$

și z_k sunt cotele standard corespunzătoare, are loc egalitatea: care arată că distanțele relative între valorile variabilei V rămân aceleași.

Inconvenientul utilizării cotelor z îl constituie faptul că acestea se exprimă prin numere cu semn și cu zecimale, fiind cuprinse, în general, în intervalul $[-3, +3]$. O rezolvare practică o constituie efectuarea unei transformări liniare asupra cotelor z , care conduce la obținerea unor cote standardizate:

$$z' = M + Sz$$

unde:

(II.5.5)

z' = cota transformată a cotei standard z ;
 M, S = media și abaterea standard pentru cotele standardizate obținute prin transformare.

În mod obișnuit, valorile z' se rotunjesc la numere întregi.

În tabelul II.5.1 sunt prezentate câteva dintre transformările utilizate mai des (Lohse, Ludwig, Röhr, 1982).

Tabelul II.5.1
 Valorile M și S pentru diverse transformări
 ale cotelor standard

	Scala							
	N	C	WP	L	T	Z	IQ	H
M	3	5	10	10	50	100	100	50
S	1	2	3	5	10	10	15	14

Cotele standardizate au aceleași proprietăți ca și cotele standard z , singura diferență între ele fiind valoarea mediei și cea a abaterii standard.

BIBLIOGRAFIE

- Aiken, L.R. (1994), *Psychological Testing and Assessment*, Allyn and Bacon, Massachusetts.
- Albu, M. (1990), *Formule pentru estimarea fidelității testelor psihologice*, "Studia Univ. Babeș-Bolyai", Psychologia-Paedagogia, 2, pag. 84-98.
- Albu, M. (1992), *O nouă metodă de interpretare a scorurilor obținute la scalele chestionarelor de personalitate*, "Studia Univ. Babeș-Bolyai", Psychologia-Paedagogia, 1-2, pag. 141-143.
- Albu, M. (1993), *Analiza de itemi: o etapă obligatorie în construirea unui test psihologic*, "Revista de psihologie", 1, pag. 75-81.
- Albu, M. (1994), *O metodă nouă de interpretare a scorurilor la scalele Inventarului Psihologic California (CPI)*, "Studia Univ. Babeș-Bolyai", Psychologia-Paedagogia, 1-2, pag. 137-141.
- Albu, M., Pitariu, H. (1991a), *O reconfigurare a noțiunii de putere de discriminare a testelor psihologice și de cunoștințe*, "Studia Univ. Babeș-Bolyai", Psychologia-Paedagogia, 1, pag. 87-92.
- Albu, M., Pitariu, H. (1991b), *Algoritm de construire a unei scale pentru un test psihologic: contribuții la re proiectarea scalei F/M a Inventarului Psihologic California (CPI)*, "Studia Univ. Babeș-Bolyai", Psychologia-Paedagogia, 2, pag. 29-35.
- Albu, M., Pitariu, H. (1992a), *Optimizarea construcției testelor psihologice și de cunoștințe*, "Revista de psihologie", 4, pag. 339-344.
- Albu, M., Pitariu, H. (1992b), *Un algoritm de construire a unei scale cu o consistență internă ridicată*, "Studia Univ. Babeș-Bolyai", Psychologia-Paedagogia, 1-2, pag. 141-143.
- Albu, M., Pitariu, H. (1993), *Proiectarea testelor de cunoștințe și examenul asistat pe calculator*, Editura Casa Cărții de Știință, Cluj-Napoca.
- Albu, M., Pitariu, H. (1994), *Redefining Criterion Validity of Psychological Tests*, "Revue Roumaine de Psychologie", 1, pag. 79-87.
- Albu, M., Pitariu, H. (1995a), *Criterion-Related Validity of Psychological Tests: A New Solution*, "European Journal of Psychological Assessment", Supplement 1, pag. 56.
- Albu, M., Pitariu, H. (1995b), *A New Method for Assessing the Stability over Time of Psychological Test Scores*, "Studia Univ. Babeș-Bolyai", Psychologia-Paedagogia, 1-2, pag. 35-45.
- Albu, M., Pitariu, H. (1995c), *O contribuție metodologică la formarea grupelor contrastante*, În: *Studii și cercetări din domeniul științelor socio-umane, prezentate la sesiunile științifice anuale 1993-1994*, Academia Română - Filiala Cluj-Napoca, Institutul de Cercetări Socio-Umare, Cluj-Napoca.

- Amelang, M., Borkenau, P. (1986), *The Trait Concept: Current Theoretical Considerations, Empirical Facts, and Implications for Personality Inventory Constructions*, În: Angleitner, A., Wiggins, J.S. (Ed.), *Personality Assessment via Questionnaires*, Springer-Verlag, Berlin, Heidelberg.
- Anastasi, A. (1954 și 1976), *Psychological Testing*, MacMillan Publishing Co., Inc., New York.
- Andrich, D. (1988), *Rasch Models for Measurement*, Sage Publications, Inc.
- Angleitner, A., John, O.P., Löhr, F.L. (1986), *It's What You Ask and How You Ask It: An Itemmetric Analysis of Personality Questionnaires*, În: Angleitner, A., Wiggins, J.S. (ed.), *Personality Assessment via Questionnaires*, Springer-Verlag, Berlin, Heidelberg.
- Arrow, K. J. (1963), *Social Choice and Individual Values*, New Haven and London, Yale University Press.
- Ausubel, D. P., Robinson, F. G. (1981), *Învățarea în școală. Introducere în psihologia pedagogică*, Editura Didactică și Pedagogică, București.
- Averill, J.R. (1980), *The Emotions*. În: Staub, E. (ed.), *Personality. Basic Aspects and Current Research*, Prentice-Hall, Englewood Cliffs, New Jersey.
- Bakeman, R., Gottman, J. M. (1986), *Observing Interaction. An Introduction to Sequential Analysis*, Cambridge University Press.
- Barclay, J.R. (1968), *Controversial Issues in Testing*, Houghton Mifflin Company, Boston.
- Baron, R.M., Kenny, D.A. (1986), *The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations*, "Journal of Personality and Social Psychology", 6, vol. 51, pag. 1173-1182.
- Benziman, H., Toder, A. (1993), *The Psychodiagnostic Experience: Reports of Subjects*, În: Nevo, B., Jäger, R. S. (ed.), *Educational and Psychological Testing: The Test Taker's Outlook*, Hogrefe & Huber Publishers.
- Blum, M. L., Naylor, J. C. (1968), *Industrial Psychology. Its Theoretical and Social Foundations*, Harper & Row Publishers.
- Boekaerts, M. (1996), *Personality and the Psychology of Learning*, "European Journal of Personality", vol. 10, 377-404.
- Brogden, H. E., Taylor, E. K., (1949), *The Dollar Criterion - Applying the Cost Accounting Concept to Criterion Construction*, Personnel Psychology.
- Burisch, M. (1986), *Methods of Personality Inventory Development- A Comparative Analysis*, În: Angleitner, A., Wiggins, J.S. (Ed.), *Personality Assessment via Questionnaires*, Springer-Verlag, Berlin, Heidelberg.
- Burisch, M. (1997), *Test Length and Validity Revisited*, "European Journal of Personality", vol. 11, pag. 303-315.
- Cascio, W. F. (1991), *Applied Psychology in Personnel Management*, Prentice-Hall, Inc., New Jersey.
- Chaplin, W.F., John, O.P., Goldberg, L.R. (1988), *Conceptions of States and Traits: Dimensional Attributes with Ideals as Prototypes*, "Journal of Personality and Social Psychology", 4, vol. 54, pag. 541-557.
- Cone, D. J. (1993), *The Current State of Behavioral Assessment*, "European Journal of Psychological Assessment", vol. 9, 3, pag. 175-181.
- Corsini, R. J. (ed.) (1994), *Encyclopedia of Psychology*, John Wiley & Sons.
- Cozby, P.C. (1993), *Methods in Behavioral Research*, Mayfield Publishing Company.
- Crehan, K. D. (1974), *Item Analysis for Teacher-made Mastery Tests*, "Journal of Educational Measurement", vol. 11, 4, pag. 255-262.
- Cresin, R. (1966), *Aplicarea indicatorului "energie informațională" în cercetarea longevității*, "Revista de statistică", 12.
- Cronbach, L. J. (1943), *On Estimates of Test Reliability*, "The Journal of Educational Psychology", 34, pag. 485-494.
- Cronbach, L. J. (1966), *Essentials of Psychological Testing*, A Harper International Edition.
- Cureton, E. (1955), *Validity*, În: Lindquist, E. F. (ed.), *Educational Measurement*, American Council on Education, Washington, D.C.
- Dahlstrom, W.G. (1993), *Small Samples, Large Consequences*, "American Psychologist", vol. 48, nr. 4, pag. 393-399.
- Davidson, O. M., Maxey, J. (1993), *Correlates of Students' Reaction to their Testing Environment*, În: Nevo, B., Jäger, R. S. (ed.), *Educational and Psychological Testing: The Test Taker's Outlook*, Hogrefe & Huber Publishers.
- Davis, F. B. (1955), *Item Selection Techniques*, În: Lindquist, E. F. (ed.), *Educational Measurement*, American Council on Education, Washington, D.C.
- De Gruijter, D. N. M., Van der Kamp, L. J. Th. (1984), *Statistical Models in Psychological and Educational Testing*, Swets & Zeitlinger, Lisse.
- Deinzer, R., Steyer, R., Eid, M., Notz, P. ș.a. (1995), *Situational Effects in Trait Assessment: the FPI, NEOFFI, and EPI Questionnaires*, "European Journal of Personality", vol. 9, pag. 1-23.

- Downs, S. (1989), *Job Sample and Trainability Tests*, în: Herriot, P. (ed.), *Assessment and Selection in Organizations. Methods and Practice for Recruitment and Appraisal*, John Wiley & Sons, Chichester.
- Ebel, R. L. (1955), *Writing the Test Item*, în: Lindquist, E. F. (ed.), *Educational Measurement*, American Council on Education, Washington, D.C.
- Edwards, A. L. (1970), *The Measurement of Personality Traits by Scales and Inventories*, Holt, Rinehart and Winston, Inc.
- Endler, N. S., Magnusson, D. (1976), *Interactional Psychology and Personality*, Hemisphere Publishing Corporation, Washington, D.C.
- English, H. B., English, A. C. (1970), *A Comprehensive Dictionary of Psychological and Psychoanalytical Terms*, David McKay Company, Inc.
- Everitt, B. S., Dunn, G. (1983), *Advanced Methods of Data Exploration and Modelling*, Heinemann Educational Books, London, Exeter.
- Eyde, L.R., Kowal, D.M. (1987), *Computerised Test Interpretation Services: Ethical and Professional Concerns Regarding U.S. Producers and Users*, "Applied Psychology: An International Review", 36 (3/4), pag. 401-417.
- Fernández-Ballesteros, R. (1993), *Behavioral Assessment: Dying, Vanishing or Still Running?*, "European Journal of Psychological Assessment", vol. 9, nr. 3, pag. 159-174.
- Fife-Schaw, C. (1995), *Questionnaire Design*, în: Breakwell, G. M., Hammond, S., Fife-Schaw, C. (ed.), *Research Methods in Psychology*, Sage Publications.
- Fisseni, H. J. (1990), *Lehrbuch der psychologischen Diagnostik*, Verlag für Psychologie, Dr. C. J. Hogrefe, Göttingen.
- Flanagan, J. C. (1955), *Units, Scores and Norms*, în: Lindquist, E. F. (ed.), *Educational Measurement*, American Council on Education, Washington, D. C.
- Fletcher, C. (1997), *The Future of Psychometric Assessment: Fostering Positive Candidate Attitude and Reactions*, "Selection & Development Review", vol. 13, nr. 4, pag. 8-11.
- Funder, D.C. (1983), *Three Issues in Predicting More of the People: A Reply to Mischel and Peake*, "Psychological Review", vol. 90, nr. 3, pag. 283-289.
- von Gilmer, H. B. (1971), *Industrial and Organizational Psychology*, McGraw-Hill Book Company.
- Goldberg, L.R. (1996), *A Public-Domain, Broad-Bandwidth, Personality Inventory Measuring the Lower-Level Facets of Several Five-Factor Models*, Comunicare prezentată la a VIII-a Conferință Europeană asupra Personalității, Ghent.

- Gough, H. G. (1989), *The New California Psychological Inventory: Implications for Research and Practice*, Lucrare dată participanților la simpozionul "Recent Advances in Psychological Assessment", Nevada, 29 aprilie 1989.
- Gough, H. G. (1991), *Scales and Combinations of Scales: What do They Tell Us, What do They Mean?* Manuscris.
- Graziano, A. M., Raulin, M. L. (1993), *Research Methods. A Process of Inquiry*, Harper Collins College Publishers.
- Gregory, R.J. (1992), *Psychological Testing. History, Principles, and Applications*, Allyn and Bacon, Massachusetts.
- Greuter, M.A., Algara, J.A. (1989), *Criterion Development and Job Analysis*, în: Herriot, P. (ed.), *Assessment and Selection in Organizations. Method and Practice for Recruitment and Appraisal*, John Wiley & Sons, Chichester.
- Grundman, N. E. (1957), *Measurement and Evaluation in Teaching*, The MacMillan Company, New York.
- Guilbert, J. J. (1987), *Ghid pedagogic pentru personalul sanitar*, Editura Medicală, București.
- Guilford, J. P. (1936), *Psychometric Methods*, McGraw-Hill Book Company, Inc., New York, London.
- Guilford, J. P. (1965), *Fundamental Statistics in Psychology and Education*, McGraw-Hill Book Company.
- Guion, R. M. (1974), *Open a New Window: Validities and Values in Psychological Measurement*, "American Psychologist", 29, pag. 287-296.
- Guion, R. M. (1986), *Changing Views for Personnel Selection Research*, Lucrare prezentată la al 21-lea Congres Internațional de Psihologie, Ierusalim, 15 iulie 1986.
- Hall, C.S., Lindzey, G. ș.a. (1985), *Introduction to Thesis of Personality*, John Wiley & Sons, New York.
- Hambleton, R.K., Swaminathan, H., Rogers, H.J. (1991), *Fundamentals of Item Response Theory*, Sage, Newbury Park.
- Hammond, S. (1995), *Using Psychometric Tests*, în: Breakwell, G. M., Hammond, S., Fife-Schaw, C. (ed.), *Research Methods in Psychology*, Sage Publications.
- Harris, W.G. (1987), *Computer-based Test Interpretations: Some Development and Application Issues*, "Applied Psychology: An International Review", 36 (3/4), pag. 237-247.
- Hendriks, A. A. J. (1997), *The Construction of the Five-Factor Personality Inventory (FFPI)*, Teză de doctorat, Groningen.

- Henry, W. E., Sanford, N. (ed.) (1972), *The California Psychological Inventory Handbook*, Jossey-Bass, Inc.
- Hettema, J., Kenrick, D.T. (1992), *Models of Person-Situation Interactions*. În: Caprara, G.V., Van Heck, G.L. (ed.), *Modern Personality Psychology*. Harvester Wheatsheaf.
- Hofstee, W. K. B., de Raad, B., Goldberg, L. R. (1992), *Integration of the Big Five and Circumplex Approaches to Trait Structure*, "Journal of Personality and Social Psychology", vol. 63, nr. 1, pag. 146-163.
- Hofstee, W.K., Smid, N.G. (1986), *Psychometric Models for Analysis of Data from Personality Questionnaires*. În: Angleitner, A., Wiggins, J.S. (ed.), *Personality Assessment via Questionnaires*, Springer-Verlag, Berlin, Heidelberg.
- Huba, G.J. (1987), *On Probabilistic Computer-based Test Interpretation and Other Expert Systems*, "Applied Psychology: An International Review", 36 (3/4), pag. 357-373.
- Huberty, C. J. (1984), *Issues in the Use and Interpretation of Discriminant Analysis*, "Psychological Bulletin", vol. 95, nr. 1, pag. 156-171.
- Jäger, R. S. (1993), *Measuring Reactions of Examiner and Examinee to Each Other and to the Psychodiagnostic Situation*. În: Nevo, B., Jäger, R. S. (ed.), *Educational and Psychological Testing: The Test Taker's Outlook*, Hogrefe & Huber Publishers.
- John, O.P. (1990), *The "Big Five" Factor Taxonomy Dimensions of Personality in the Natural Language and in Questionnaires*. În: Pervin, L. A. (ed.), *Handbook of Personality: Theory and Research*, The Guilford Press, New York.
- Johnson, J.A. (1996), *Predicting Observers' Ratings of the Big Five from Self-Report Inventories*, Comunicare prezentată la a VIII-a Conferință Europeană asupra Personalității, Ghent.
- Klausnitzer, J.E. (1992), *Les tests d'intelligence*, Marabout, Alleur.
- Kline, P. (1993), *The Handbook of Psychological Testing*, Routledge, New York.
- Lafon, R. (1973), *Vocabulaire de psychopédagogie et de psychiatrie de l'enfant*, Presses Universitaires de France, Paris.
- Landau, E. (1979), *Psihologia creativității*, Editura Didactică și Pedagogică, București.
- de Landsheere, G. (1975), *Evaluarea continuă a elevilor și examenele. Manual de docimologie*, Editura Didactică și Pedagogică, București.
- Landy, F. L. (1985), *Psychology of Work Behaviour*, The Dorsey Press Chicago, Illinois.
- Landy, F. L. (1986), *Stamp Collecting Versus Science. Validation as Hypothesis Testing*, "American Psychologist", vol. 41, nr. 11, pag. 1183-1192.

- Landy, F. J., Rastegary, H., Thayer, J. F., Colvin, C. (1991), *Time Urgency: The Construct and its Measurement*, "Journal of Applied Psychology", 76, 5, pag. 644-657.
- Lay, C.H. (1997), *Explaining Low-Order Traits Through Higher-Order Factors: the Case of Trait Procrastination, Conscientiousness, and the Specificity Dilemma*, "European Journal of Personality", vol. 11, pag. 267-278.
- Li, H., Rosenthal, R., Rubin, D.B. (1996), *Reliability of Measurement in Psychology: From Spearman-Brown to Maximal Reliability*, "Psychological Methods", vol. 1, nr. 1, pag. 98-107.
- Lindeman, R. H. (1978), *Evaluarea în procesul de instruire*, În: Davitz, J. R., Ball, S. (ed.), *Psihologia procesului educațional*, Editura Didactică și Pedagogică, București.
- Lindvall, C. M. (1967), *Measuring Pupil Achievement and Aptitude*, Harcourt, Brace & World, Inc., New York, Chicago, San Francisco, Atlanta.
- Lohse, H., Ludwig, R., Röhr, M. (1982), *Statistische Verfahren für Psychologen, Pädagogen und Soziologen*, Volk und Wissen Volkseigener Verlag Berlin.
- Lord, F. M. (1980), *Applications of Item Response Theory to Practical Testing Problems*, Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey.
- Lord, F. M., Novick, M. R. (1968), *Statistical Theories of Mental Test Scores*, Addison-Wesley Publishing Company.
- MacCallum, R.C., Mar, C.M. (1995), *Distinguishing Between Moderator and Quadratic Effects in Multiple Regression*, "Psychological Bulletin", vol. 118, nr. 3, pag. 405-421.
- Marcus, S. (1997), *Empatie și personalitate*, Editura ATOS, București.
- McCrae, R., Costa Jr., P.T. (1995), *Trait Explanations in Personality Psychology*, "European Journal of Personality", vol. 9, pag. 231-252.
- Megargee, E. J. (1972), *The California Psychological Inventory Handbook*, Jossey-Bass Inc., Publishers, San Francisco, Washington, London.
- Messick, S. (1995), *Validity of Psychological Assessment*, "American Psychologist", vol. 50, nr. 9, pag. 741-749.
- Meuris, G. (1985), *La méthode des tests... à la croisée des chemins*, "Bulletin of the International Test Commission", 21, pag. 5-12.
- Michell, J. (1997), *Quantitative Science and the Definition of Measurement in Psychology*, "British Journal of Psychology", 88, pag. 355-383.
- Minulescu, M. (1996), *Chestionarele de personalitate în evaluarea psihologică*, Garell Publishing House, București.